Vol. XI Part II The British Journal of Statistical Psychology November 1958

SCIENTIFIC METHOD IN PSYCHOLOGY: I

By T. R. Willis
University of Birmingham

I. *The Prevailing Confusion.* II. *Aristotle versus Galileo* III. *The Need for Causal Laws.* IV. *The Hypothetico-deductive Procedure.* V. *Summary and Conclusions.*

I. The Prevailing Confusion

Divergent Views. The conflicting pronouncements on scientific methodology, which have recently appeared in this and other psychological journals have, if anything, only increased the confusion that reigns in the minds of the ordinary research worker who has to apply the methodology. The issue over which various contributors chiefly disagree seems to be one of major importance—namely, what is the general mode of proof to be adopted in a branch of science which, like psychology, is concerned with highly complex phenomena. Dr. Richardson and Dr. Eysenck advocate one type of procedure—though for widely different reasons: their procedure has been severely censured by Gregory; and Gregory's arguments are criticized by Professor Burt, who nevertheless agrees with neither Eysenck nor Richardson (this *Journal*, V, pp. 208f., VI, pp. 44f., 107f.). In the statistical field Burt and Thurstone favour a methodology that stems from Galton, Pearson, and R. A. Fisher; Hogben, on the other hand, rejects all attempts to apply correlational, factorial, or multivariate analysis to social and psychological problems[1]. Professor Hogben is supported by Dr. J. G. Taylor who, very opportunely, reminds us of Lewin's " familiar warning against applying in psychology the Aristotelian notion that frequency is equivalent to law ". Taylor, like Burt, goes to the physicists for his model of scientific method, but draws quite the opposite conclusion. He flatly denies that the psychologist's hypotheses are statistical hypotheses and that the only test the psychologist can apply is a statistical test, as Burt had recently maintained (this *Journal*, XI, pp. 67f.); and contends that in psychology the only worthwhile experiments are those designed to test " hypotheses that have some pretensions to being accepted as universal laws of nature " (*Brit. J. Psychol.*, XLIX, 1958, pp. 106f.).

Similar divergences of view emerged during the ' Symposium on Causal Laws in Psychology ' (C. A. Mace, B. A. Farrell, and M. M. Braithwaite, *Proc. Arist. Soc. Supp.*, XXIII, pp. 31–68) and at the International Congress of 1957. It is therefore hardly surprising if Mr. Pilkington, after subjecting all these forms of ' dogmatic methodism ' to drastic criticism, should conclude that

[1] L. Hogben, *Statistical Theory* (1957). Hogben's criticisms of Burt and Thurstone will be found in his chapter on " The Impasse of Factor Analysis "; and his objections to Fisher's techniques as applied to problems of psychology and para-psychology in his concluding chapter.

there is, and can be, *no* " correct scientific procedure for the psychologist ", and that " the prescriptions laid down on supposedly methodological grounds are nothing more than a pseudo-methodology " (*Bulletin*, XXXIV, 1958, p. 62).

The Historical Approach. In spite of all this, it is difficult for those engaged in active research to avoid a feeling that, like Hogben, Mr. Pilkington has carried scepticism too far, and that, between them, they empty out the baby of quantitative psychology with the bath-water of statistical technique. Pilkington's wholly negative conclusions surely fly in the face of the views expressed by nearly every writer on scientific method, from Jevons to Stebbing and Braithwaite. The philosophers at any rate seem agreed that, when we survey the different realms of natural science, from Galileian physics to modern biology and physiology, we find " a general logical pattern common to them all "; and with them I would suggest that the best clue to problems of methodology is to be found, not in logical considerations, as Professor Burt assumes, but in historical. After all, induction itself (on which all empirical science must rest) is incapable of logical justification: we accept it solely because the history of science shows that it works.

II. Aristotle versus Galileo

Galileo's Rejection of Greek Methodology. Modern science begins, not in the age of the Greek debaters, but as one of the revolutionary movements of the seventeenth century and the daring conception of a single genius. " Scientific method as we understand it today," says Russell, " comes into the world full-fledged with Galileo ". As almost every schoolboy learns, the reform that Galileo effected in traditional methodology was as swift and sweeping as that of Luther when he overthrew the domination of the Pope, or Cromwell when he put an end to the autocracy of the Stuarts. Aristotle was dethroned[1].

The clearest and most succinct account of the ' Galileian technique ' is that given by Professor Stebbing (*A Modern Introduction to Logic*, pt. ii, pp. 490f., cf. pp. 299f.). " Galileo," she writes, " first formulates a hypothesis in mathematical form; then deduces the consequences; and finally, tests it experimentally " (pp. 490–1). However, most of her illustrations are drawn from the work of Galileo and his followers in the field of astronomy; and the practical psychologist will no doubt immediately ask: " is it fair to take a physical science like astronomy as our model when we are really concerned with problems of consciousness or mind? " This question Professor Stebbing herself takes up on a later page, and counters with another: " Can psychology be regarded as a science in the same sense as the physical sciences? " Her answer is: " It is the method used that determines whether or not a given study is in fact a science ".

The view I wish to support is in line with hers; and in the field of psychology has been perhaps most clearly argued by Lewin in his chapter on " The Conflict

[1] In their replies to Dr. Richardson, both Professor Burt (who rejects Russell's dictum, quoted above) and Mr. Gregory seem eager to disparage the highly original contributions of Galileo. But the only historical evidence they offer is a vague allusion to certain unspecified ' anticipations ',

between Aristotelian and Galileian Modes of Thought in Contemporary Psychology" (*A Dynamic Theory of Personality*, 1935), to which Dr. Taylor briefly alludes[1]. Lewin accepts the view that scientific method is the same in every branch, and contends that it may be seen at its best in Galileian physics. The practical moral which he draws is that, if psychology is to become a branch of natural science instead of a relic of ancient metaphysics, it must abandon its traditional devotion to Aristotelian concepts and modes of proof, and adopt instead those of Galileo.

Aristotle treated all sciences as essentially deductive. His procedure was based on the assumption of heterogeneous ' substances ', i.e. physical or terrestrial bodies and metaphysical or celestial bodies (including ' souls '), each existing independently of the rest. The various ' substances ' are classified in a hierarchical scheme according to their qualitative ' attributes ' and qualitative ' states '; and propositions in science are regarded essentially as predictions about substances and their attributes, while the contributions of mathematics are entirely ignored. When a Greek philosopher reasoned about nature, his reasoning was purely syllogistic; and his explanations of observable phenomena were always sought in the ' properties ' which the various ' substances ' possessed—e.g. the ' propensities ' of material bodies to rise or fall or of the mind or soul to perceive, reason, and strive—i.e. to its inherent ' faculties ' and its ' inner drives '. In short, as Whitehead puts it, " the Aristotelians said *classify* when they should have said *measure* " (*Science and the Modern World*, p. 66).

Of substances, whether material or mental, Galileo knows nothing. For qualities he substitutes continuously varying quantities; and for the fictitious ' inner drives ' he substitutes ' directed forces ', arising from the ' environmental situation ' and ' acting in accordance with causal laws '. These laws, and those of Newton (which are really a development of Galileo's) are ' scientific laws ' in the true sense, not mere ' statistical laws '; they state what does and must hold universally, not what probably holds in a majority of cases. Nevertheless, as Dr. Taylor adds, " in psychology Aristotelian modes of thought are today as prevalent as ever, backed up by an elaborate structure of statistical procedures ".

The Characteristics of Modern Sciences. Professor Burt's endeavour to outline a methodology which will justify the statistical approach (this *Journal*, XI, 1958, pp. 31f.) seems to overlook the distinctive features of modern scientific procedure. I suggest that they may be reduced to four—all of them due to the versatile genius of Galileo. (i) To begin with, Galileo's substitution of quantitative variation for crude differences of quality implied the use of exact *measurement* in place of mere classification according to ' nature ' or ' kind '. His proofs are always mathematical, and not syllogistic. (ii) Still more important was his introduction of the notion of *experiment*. Instead of arguing *a priori* that a ball of iron ' must ' fall to the ground with a greater speed than a ball of wood because

[1] Here I should like to say that my own paper was in fact submitted to the editor just before Dr. Taylor's article appeared in print, and the editor has kindly allowed me to modify my original draft to deal with some of the points there raised.

iron is ' heavier ', he drops iron and wooden balls from the tower of Pisa, and shows they they reach the ground simultaneously. (iii) Thirdly, and largely as a consequence of the other two, we have the introduction of scientific *instruments*, typified by Galileo's invention of the telescope—the forerunner of many other types of *ad hoc* apparatus designed to convert observation into experiment and to increase the precision of the measurements received. (iv) Finally, he demonstrated what may be called *the homogeneity of Nature*. Instead of a heterogeneous collection of disparate beings, some ruled by law, others by chance or their own arbitrary wills, he reveals to us a strictly deterministic universe, everywhere conforming to intelligible causal laws. Thus, on turning his telescope on to the various objects in the sky, he discovered that the sun, the most impressive of all celestial bodies, nevertheless exhibited darkened spots, like the shadows on the moon, which, when their speed is duly measured, enabled him to calculate the exact speed of the sun's rotation, and so finally to draw the manifest conclusion: the sun and stars are not, as Aristotle supposed, immortal and immutable entities, composed of semi-divine ' essence ', but matter of precisely the same kind, in precisely the same states, and precisely obeying the same equations of motion, as the matter composing the objects familiar to us on the earth. Moreover, the ' heavenly bodies ' residing in ' free space ' do not (as the Greeks maintained) move in ' perfect circles ' because they themselves are perfect: they move, like rolling balls, in simple straight lines when unimpeded, and, when subject to external forces, in closed ellipses, or those elongated ellipses we term parabolas, like projectiles or drops from a fountain. In short, as Lewin expresses it, " the same law rules the courses of the stars as rules the falling of stones; and thus the outlook of a Galileo is determined by the idea of a comprehensive unity of the whole physical universe ". Man himself, and his behaviour, form just another set of similar phenomena in the self-same physical universe (Lewin, *loc. cit.*, p. 10).

Psychology, therefore, to become genuinely scientific, must adopt the same approach. It must rely first and foremost on measurement. It must employ planned experiment instead of intuitive ' observation ' and clinical ' insight '. For this it requires the instruments and the apparatus of a modern laboratory. And, last but not least, it must abandon the notion of semi-divine substances, with occult ' powers ' or ' faculties ', such as the ' souls ', ' minds ', or the conscious or unconscious ' egos ' of the traditional jargon, and must recognize the unity of all observable phenomena within a single physical world.

Scientific Laws versus Statistical Laws. As a corollary to all this, so Lewin has argued, we must discard the pseudo-scientific devices of modern psychological statistics. " The statistical procedure, at least in its commonest application in psychology, is the most striking expression of an outdated Aristotelian mode of thinking: the classificatory character of its concepts is sufficient to show that it is still essentially qualitative under a thin mathematical disguise." (Lewin, *loc. cit.*, p. 16.) Lewin's warning fell on deaf ears; and Dr. Taylor's restatement of his arguments in the context of more recent psychological

studies are most opportune. As he rightly points out, the growing preoccupation with statistical devices and Fisher's so-called ' rules of experimental design' have proved to be, in psychology at any rate, " nothing but a cloak for intellectual sterility ". " What psychology needs is scientific laws not statistical laws and such laws are unlikely to be found by statistical investigations; and the kind of answers we must look for are not statements of probability, but general laws applicable throughout the whole field of psychology."

III. The Need for Causal Laws

The Alleged Implications of the Quantum Theory. In his paper on ' Quantum Theory and the Principle of Indeterminacy ' Professor Burt seeks to rehabilitate the statistical approach by invoking the probabilistic hypotheses advanced by later quantum theorists, supplemented by a formidable mathematical apparatus of Fourier analysis, Fourier integrals, complex numbers, matrix algebra, and the like. He thus seems to live in a double world—a world of law (described in his first paper[1]) and a world of chance (described in his second paper), and tries to make the best of both. Conscious processes, it appears, like quantum processes, are to be assigned to the world of chance or at least of voluntary caprice. The reasons are left obscure. Apparently they consist in certain ' formal analogies ' described as ' often striking '. But what precisely the analogies are which Professor Burt thus discovers, or thinks he can discover, between quantum phenomena and mental phenomena respectively, is never explicitly stated. The conclusion seems to be that " the brain or nervous system can no longer be treated as a ' determinate machine ' ". Presumably we are to think of it as an ' indeterminate machine ', like a roulette-wheel or a dice-box. Yet surely even the fall of the dice and the behaviour of the wheel are strictly determined by causal forces: they seem ' indeterminate ' to us simply because they are too numerous or minute for us to observe or determine. If, for instance, I say: " Professor Burt is probably at this moment sitting down at lunch ", the word ' probably ' merely implies a doubt due to my personal ignorance; actually Burt either is or is not at lunch: the fact itself has nothing ' probabilistic ' about it. Similarly with every other object from the electron to the furthest star.

Scientific Laws as Causal Laws. The Galileian view of physics rests implicitly on a doctrine of universal causation, and thus eliminates the antithesis between events that are due to chance and events that are governed by law. The laws of the physical sciences are expressed by the various equations of dynamics, which state in quantitative form the precise relations between cause and effect. Since, therefore, wherever a complete analysis is possible, science always reveals strict causal determination, psychology, if it is to rank as a genuine science, must also aim at the discovery of exact and universal causal laws. Its

[1] In this first paper he cites, as Dr. Taylor has done, Newton's equation for the force of gravity as a typical law of physics; but, whereas Dr. Taylor describes it as exact and universal, Professor Burt describes it as a mere ' hypothesis ', and so presumably would regard it as merely a ' probable ' law.

hypotheses must be causal hypotheses; and the causes must be real causes. They must be concrete events or entities, continuous with their effects in *actual* space and time, not abstract ' factors ', resulting from a mere mathematical analysis, and lying in a mere ' configurational ' space of the Thurstonian type. Thomson warns the statistical psychologist never to ' reify ' his factors, or imagine that a ' man has a *g* just as he has a liver '; we would rather warn him (as Miles has done[1]) to reject any factors that are *not* real. After all, the aim of the scientist is to depict the world as it really is. A man must have a brain; and, if the statistical psychologist is never to look for the causes of behaviour in the concrete state of the individual's brain or nervous system, he has ceased to be a psychologist and become a pure statistician.

Consider the regular type of question that psychology seeks to answer. What kind of stimuli cause this or that sensation or behavioural response? What makes us forget when we ought to remember? What are the causes of backwardness, delinquency, and psychoneuroses, and what types of treatment will prove causally efficacious? And so on. All these are questions of causal type. And Russell himself has assured us that an essential postulate of all scientific inferences is the assumption of ' causal lines ' (*Human Knowledge*, pp. 333f., 471f.).

IV. The Hypothetico-Deductive Procedure

The Standard Practice. Galileo's great contribution to scientific method was the invention of the hypothetico-deductive procedure—" testing deduction from hypotheses by comparison with facts—the sovereign rule of science " (Stebbing, *loc. cit. sup.*). Except for statistical psychologists of the Fisher school, this basic procedure is endorsed by nearly every psychological writer. Open almost any standard textbook, say *The Foundation of Psychology* (Boring, Langfeld and Weld): there in the very first chapter we read that the " safe rule for research is that commonly known as the hypothetico-deductive method ". Mr. Pilkington associates the method specifically with Hull and Eysenck. But among the other groups that he names, each of the writers (with the possible exception of McDougall and Allport) seems also to have adopted it.

Mr. Pilkington himself apparently holds that there is " no one eternally correct method " which applies to psychology in common with other sciences; and he quotes Peters as deprecating the attempt to deduce from the physical sciences " prescriptions laying down, on supposedly methodological grounds, the course psychology ought to pursue to achieve success ". To this there is a two-fold answer: (i) to discover what type of methodology is most likely to ' achieve success ', the safest place to look is surely those sciences where success has admittedly been achieved, namely, the physical; (ii) before we decide that this type of methodology is unsuitable for psychology and start looking around for something different, the proper procedure is to try it first of all, and see if it really does ' achieve success '.

[1] Miles, T. R., ' On Defining Intelligence ', *Brit. J. Psychol.*, XXVII, 1957, pp. 160f.

A Concrete Example. Accordingly, as a recent and typical illustration of the success with which the hypothetico-deductive method may be applied to psychological problems, may I cite the research on patients in a mental hospital just published by Dr. Hildebrand (*Brit. J. Psychol.*, XLIX, 1958, pp. 1–11)? The object was " to test Jung's hypothesis ", as reformulated by Eysenck, that hysterical and anxiety neuroses result from extreme degrees of extraversion or introversion superimposed on a general factor of neuroticism. I choose this inquiry partly because it is largely statistical and to that extent in conformity with the requirements of Gregory and Burt, and partly because the hypothesis which it establishes happens to be the one criticized by Gregory himself in a previous issue of this *Journal* (V, pp. 208–212): if therefore he still rejects it, what other alternative theory can he suggest in its place?

Dr. Hildebrand begins by stating his hypothesis in the form of three main propositions; five corollaries are deduced; and these are then subjected to an experimental test. With one minor exception, fully accounted for by a subsidiary hypothesis, " all the deductions are supported at a high degree of significance "; and we thus obtain a battery of verified tests which should enable the clinician to divide a mixed population into " three categories in accordance with the original hypothesis ", viz., neurotics of the extraverted and the introverted types and the normal—a most valuable result.

Surely, then a methodology which has satisfied scientists from Galileo to Russell, and serves psychology so well, should provide an adequate basis for present-day investigators. Nevertheless, in the papers cited both Mr. Gregory and Professor Burt seem to reject all this. We hear a good deal about " the *a priori* probabilities of simple and complex hypotheses ", and Mr. Gregory gives a statistical formula for calculating the *a posteriori* probability; but what precisely is meant by ' the paradox of the simplicity postulate ', and how the formula proposed emerges from Mr. Gregory's theory, the reader is never told.

In his earlier writings Professor Burt himself seems to have held the view that I am here defending. He then maintained that the logical form of inductive proof was the same for all empirical sciences, from the simplest to the most complex. His chief complaint was that psychological hypotheses, particularly in factorial research, were " far too static ", and he registered a strong plea for what he called " psychodynamics ". The logical models he proposed were avowedly based on those of ordinary dynamics; and, as Mr. Jameson has pointed out, in *The Factors of the Mind* " he came forward as one of the earliest advocates of the hypothetico-deductive method in psychology " (see chap. II), and even argued that ' factors ' might be capable of supplying ' causal explanations '.

However, in the recent discussions of the subject (*Educational Review*, X, 1958, pp. 118f.), as well as in his contributions to this *Journal*, he appears to have renounced his former views. Like Mr. Gregory, he seeks to disparage Galileo's contributions to methodology, and prefers to fall back on the older logical techniques of Aristotle and Archimedes. This apparently leads him to trust in what he calls " the consilience of inductions "—a phrase which he leaves

unexplained. The hypothetico-deductive procedure he now dismisses as fallacious. But where precisely does the fallacy lie? He is equally emphatic in his strictures on the concept of causation and determination in psychology which go with it. " Today ", he declares, " this methodological standpoint appears quite anachronistic"—surely a gross overstatement.

I fully agree that the methods of the clinical psychologist and the industrial psychologist have been lacking both in rigour and in self-consistency. But they have to adjust their procedures to the practical exigencies of their work. However, psychology, as such, is or professes to be a theoretical science; and it is with theoretical sciences (such at least as are ' inductive ') that I am here alone concerned.

V. Summary and Conclusions

May I therefore suggest that, in spite of what Mr. Pilkington has told us, there *is* " a logic of explaining behaviour ", which is not perhaps " finally settled", but accepted in principle by most present-day psychologists? And, if I am asked what precisely is the method to be followed by the research student in his studies of behaviouristic problems I could perhaps best sum up my answer—what Mr. Pilkington would call my " prescriptions for the sort of course psychology should pursue in order to achieve success "—in the following terms. (1) The procedure most suited for psychological inquiries is the experimental method introduced into modern science by Galileo; (2) this procedure implies (in the words of Mr. Pilkington) that " the ideal of scientific explanation " is one best expressed " in terms of the hypothetico-deductive system ", i.e. by a method that seeks to verify laws by determining whether or not the facts are consistent with the law hypothesized; (3) these laws are, or should be, causal laws, i.e. the psychologist should seek to prove laws stating the verifiable causes of the observable events.

Vol. XI Part II The British Journal of Statistical Psychology November 1958

SCIENTIFIC METHOD IN PSYCHOLOGY: II

By Cyril Burt and W. L. Gregory
University College, London

I. *Galileian versus Aristotelian Methodology.* II. *The Hypothetico-Deductive Method.* III. *The Demand for Causal Laws.*

I. Galileian versus Aristotelian Methodology

The Origins of Modern Scientific Procedures. The criticisms which Mr. Willis has urged against the papers published by one or other of us in this journal turn on questions about which we hold much the same views. Hence, to avoid repetition, we have amalgamated our answers into a single joint reply. Mr. Willis, at the end of his article, summarizes his arguments under three main heads. It will therefore be convenient to adopt that arrangement here. His first contention is that "the procedure most suited for psychological inquiries is the experimental method introduced by Galileo".

Since Lewin wrote his celebrated chapter, it has become the fashion among certain psychologists to claim the prestige of Galileo for the particular scientific methods they themselves favour. We are tempted to echo the punning text on which Galileo's opponent, Fra Caccini, preached his celebrated sermon: "Ye men of Galileo, why stand ye gazing up into the heavens?" The fact that it was Galileo who effectively revolutionized the "theory of the heavenly bodies" which mediaeval astronomers had inherited from Hipparchus and Ptolemy is no reason why a "return to Galileo's concepts and modes of proof" should help the modern psychologist to cope with the peculiar problems presented by human behaviour. However, since the history of scientific method has aroused such widespread interest and is so often misrepresented, it will be well to begin by briefly reviewing the relevant facts.

In the first place, it is important to note that Lewin's account of the origins of scientific method, which both Mr. Willis and Dr. Taylor implicitly adopt, would be accepted by no contemporary historian. The notion that Aristotle never got beyond the syllogism, and that the pre-Galileian schoolmen never got beyond Aristotle, contradicts the recorded facts. So far as Aristotle is concerned, Lewin and Mr. Willis seem merely to have dipped into his earliest works—the *Physics* and *The Heavens* (both written while Aristotle was still dominated by the teleological doctrines of the Socratic school); *The Posterior* (i.e. 'later') *Analytics* and his pioneer researches in biology are entirely overlooked. Aristotle was the son of a physician, and received his training in biology and medicine, not in astronomy. It is therefore hardly fair to judge him exclusively by the

latter[1]. Here he merely adopted the geometrical theory of concentric and geometrical spheres developed by Eudoxus (like Aristotle, a pupil of Plato), and, in his matter of fact way, transformed what with Eudoxus was merely a mathematical hypothesis into a mechanical one—quite in the spirit of Mr. Willis himself.

As regards methodology there is no real antagonism between Galileo and Aristotle. When Galileo cites Aristotle's views on method, it is almost always as those of an authority whom he and his critics revere—as "il primo, unico, ed ammirabile esplicator della dimostrazione, degli elenchi, ed in somma di tutta la logica". What he castigates are the fanciful theories in physics and astronomy held by certain "Aristotelians"; and by "Aristotelians" he means theorists of the Dominican school, who, like their master, St. Thomas, regarded Aristotle as the final authority on *all* problems of natural science, including physics and astronomy—*il maestro di color chi sanno*[2].

The Historical Approach. We should fully agree with Mr. Willis when he maintains that the value of the various scientific techniques that together make up our present-day methodology is to be estimated, not by purely logical considerations, but in the light of their past history—the philosophical background that suggested them and the type of practical problems they were designed to meet. But we cannot help feeling that his own account of scientific history is as faulty as that of Lewin. The progress of scientific method, like that of political institutions to which he compares it, is the cumulative result of many contributory minds, not the achievement of 'a single genius'. Mr. Willis complains that we offer no 'historical evidence' for what he considers to be our 'disparagement' of Galileo's highly original contributions. May we therefore try, once for all, to dispel the popular notion of a 'Galileian revolution' which overthrew the 'autocratic rule' of Aristotle and his followers?

[1] Those who quote Lewin's criticisms commonly stress two main points, neither of which is strictly true. (i) "Aristotle," we are told, "regards propositions in science as essentially statements about substances and their qualities, and ignores the contributions of mathematics". As a matter of fact, Aristotle enters at some length into the relation of mathematics to physics: he is especially concerned about the conditions that render some sciences "prior to and more exact than others". His conclusion resembles Plato's—that mathematics owes its 'priority' and its 'exactitude' to its abstract character, and for this reason is "basic to physics, but less complete": "the mathematician investigates abstractions only, for he strips off qualities and leaves only the quantitative and the continuous" (*Metaph.*, K, iv, 1061 a 30). (ii) "Aristotle", it is said, "treats all sciences as essentially deductive". This may be the impression gained if one gets no further than the account of syllogistic reasoning in the *Prior Analytics*. But in his later writings he emphasizes that, while mathematics is deductive and deals with what is 'necessary', the generalizations which natural science tries to reach "are got by induction from sense-perception". Hence, though mathematics cannot ascertain causes, it can supply *reasons* for the *facts* studied by the quantitative sciences—optics, harmonics, and astronomy—and at times even by non-quantitative sciences, such as medicine, which are not directly 'subaltern' to it. (For the quotations and references, see W. D. Ross, *Aristotle*, *s.v.*; Lewin unfortunately gives neither quotations nor references to support his account.)

[2] See the amusing passage in the *Dialogo dei Massimi Sistemi* (*Giorn.* Ia) beginning: "Di grazia, signor Salviati, parlate con piu rispetto d'Aristotile". As Galileo's supporter goes on to explain, "one may be a grand expert on logical method and yet not be so expert in applying that method to practical problems".

(*a*) *General Methodology.* So far as methodology is concerned, there is little or nothing that is entirely 'original' in Galileo's work. The theory of it is to be found in Aristotle; and the experimental method was discussed and practised by his followers both in the Lyceum and at Alexandria, notably (as we know from Galen) by the medical school. After the Dark Ages its revival was due largely to the experimental work of the Franciscan School founded at Oxford by Grosseteste (1168–1253, Bishop of the Diocese and the first Chancellor of the University)—a school which included such celebrated names as those of Roger Bacon, Duns Scotus, and William of Ockham. The Franciscans, unlike the Dominicans, preserved the earlier Platonic and Neo-Platonic tradition, which from the time of St. Augustine to that of St. Thomas Aquinas had furnished the basis of Catholic philosophy. One of its cardinal features was the value of a mathematical approach[1].

(*b*) *Specific Techniques.* Mr. Willis singles out four main "characteristics of modern science, ... all due to the bold originality of Galileo". To examine this fourfold claim will involve a rather lengthy digression. But, if what we say induces those who are interested in the problems of methodology to glance at the original sources, they will find their efforts amply rewarded; the development of method in the physical sciences while they were new is full of suggestive parallels for the development of method in the new science of psychology.

(i) Mr. Willis tells us that Galileo was the first to substitute systematic measurement for mere qualitative classification. Galileo's proofs, he adds, "are always mathematical, not syllogistic". And he exhorts the psychologist

[1] Much of the progress no doubt was due to the advances in technology and other crafts. But this was by no means the sole reason for the re-awakening of scientific interests at the close of the Dark Ages. In 1120 a copy of Euclid's *Elements* fell into the hands of Adelard of Bath, who promptly published a Latin translation. And during the next half century Aristotle's *Posterior Analytics*, a treatise on *Mechanica Problemata* supposed to be his, the writings of Archimedes on mechanical and hydrostatical problems, and the work of Apollonius on *Conic Sections*, all became accessible. Their impact produced a rapid development of what soon became known as the *logica nova*, in contrast to the traditional *logica vetus*—the deductive logic of substance and quality, based on the syllogism, as expounded in the earlier pages of the *Prior Analytics*. Above all it was the methodology of the *Posterior Analytics* (on which Grosseteste wrote a lengthy commentary), far more than Aristotle's views on deduction or on metaphysics, that excited the interest of the Oxford school during the next two hundred years. However, with the death of Ockham (1349)—the last of the great scholastics—original work on scientific problems in Britain lapsed almost entirely, until it was revived a couple of centuries later by Gilbert, Harvey and Newton. And during the 15th and 16th centuries the lead was taken over by the Paduan school.

Padua had already sent Paulus Venetus to Oxford to study the techniques of the English scientists; and Galileo's own views on the subject seem to have been largely borrowed from Zabarella's treatise, *De Methodis*, which frequently alludes to Grosseteste's contributions. Galileo himself was plainly familiar with the work of the British scholastics. When he refers to "*un verissimo assioma d'Aristotile che i' insegna che Frustra fit per plura quod potest fieri per pauciora*" ('Ockham's razor'), he is citing *verbatim* the maxim as given by Duns Scotus and later quoted by Ockham; and, in discussing attraction by gravity, he discourses at length on the analogy with magnetic attraction suggested by Gilbert (*De Magnete*, 1600) whose book, he says, was presented to him by *un filosofo peripatetico*. We would suggest that, if any such contrast as Lewin describes is relevant to our problem, it should be the contrast between what the 13th century writers had called the 'old logic' and the 'new'—the logic of science which sought to combine the *via rationis* with the *via experimenti*.

to follow suit. Now, in Galileo's writings there are, as a matter of fact, surprisingly few measurements; and, when he insists on the importance of 'mathematical' proofs, he is thinking not so much of a quantitative treatment as of a geometrical. As he himself explains, his 'mathematical' demonstrations are modelled on those of Euclid and Archimedes. It was not Galileo but Tycho (*d.* 1601) who introduced the new standard of accurate measurement into astronomy, and improved the traditional instruments for this purpose. Tycho too was the first to adopt the principle of repeating and combining quantitative observations for each datum, so that he was able to reduce the errors in the longitudes and latitudes of the fixed stars to 25″ or less[1]. Nor do we encounter anything in Galileo's work comparable to Kepler's mathematical demonstrations (1609) of the laws of planetary motion on the basis of Tycho's measurements of their coordinates.

(ii) It is quite mistaken to think of Galileo as primarily an experimental scientist. Most of his experiments are thought-experiments, treated rather as a device for convincing the sceptic than as an essential method for discovery or verification. Were we to single out, from among the leading scientists of this particular period, the most typical experimentalist, we should choose Gilbert of Colchester, whose treatise on magnetism and frictional electricity (*De Magnate: Physiologa Nova*, 1600) is, as a modern commentator observes, "characterized almost throughout by its reliance on the results of experiment"[2].

(iii) Galileo did not invent nearly as many scientific instruments as Tycho; and he certainly did not invent the telescope: a better claim could be made out for yet another Englishman, Thomas Harriot, who records that in 1585 he took with him a telescope of his own invention on his voyage to the newly discovered land of Virginia[3]. Nor was Galileo the earliest to observe sunspots. Chinese astronomers had noticed them centuries earlier. Kepler studied one with a pin-hole camera, supposing it to be Mercury in transit. The man who contributed most to their investigation was Father Scheiner—Professor of Mathematics

[1] The notion of measuring the relative positions of stars in terms of their coordinates was in fact due to one of the much despised Greeks—Hipparchus (*c.* 130 B.C.), who invented the astrolabe (the oldest scientific instrument) for that purpose. Anyone who takes the trouble to compare his specifications, as given in Ptolemy's catalogue (*Almagest*, Bk. VII), with those of a modern star atlas will be amazed at their precision, and especially at the accuracy of his calculated estimate of the mutation of the earth's axis.

[2] Cf. E. Zilsel, 'The Origins of Gilbert's Scientific Method', *J. Hist. Ideas*, II, pp. 1f. Incidentally, Galileo did not himself drop balls from the tower of Pisa, as Mr. Willis supposes. On the contrary, his discussion makes it clear that he is referring to experiments already carried out. The first to climb the tower for such a purpose was in fact an Aristotelian (G. Coresio, 1612: see this *Journal*, XI, p. 47, footnote 1). The experiment on rolling balls down inclined planes had also been carried out much earlier by Jordanus Nemerarius (*De Ratione Ponderis*, *c.* 1225) and repeated by Leonardo (who had noted their uniform acceleration).

[3] S. P. Rigaud, *Supplement to Bradley's Miscellaneous Works with an account of Harriot's Astronomical Papers*, Oxford, 1833, pp. 17f. Harriot was Sir Walter Raleigh's mathematical adviser, and a member of the astronomical 'school of night' burlesqued in *Love's Labour's Lost*. Roger Bacon (*Opus Maius*, *c.* 1250), however, had already described how by means of lenses it was possible to "make distant objects appear larger as if near"; and Thomas Digges claims that his father Leonard (another Oxford mathematician), after reading 'Bakon's Experiments', had constructed a 'Perspective Glass', i.e. a telescope without a tube (*Stratioticos*, 1578, p. 189).

at Ingolstadt, who was apparently the first to construct an 'astronomical' telescope and place it on an equatorial mounting. Those who imagine Galileo's work to be a model of the 'new scientific methodology' would do well to compare his writings (chiefly witty polemical dialogues, written in delightful Italian, and addressed not so much to the scientific specialist as to the educated public) with Scheiner's Latin tome of nearly 800 pages, *Rosa Ursina sive Sol*, a systematic treatise which records and analyses nearly 2000 observations of sunspots (with many graphs and illustrations), and forms a model of scientific method[1]. Though undoubtedly an 'Aristotelian'[2], so far from believing that the sun consists of a 'celestial and immutable substance' (as Mr. Willis imagines all Aristotelians did), he argues, from his own observations as well as from accredited authorities, that the sun is after all *ex natura sua corruptibile* and of the same elements as the earth.

(iv) This brings us to the last of the four contributions to methodology which Mr. Willis says we owe to Galileo, namely, "the idea of a comprehensive unity of the physical universe". If that merely means it is a sound working principle to start with familiar and well-established concepts before invoking entirely new ones, we should wholly agree. Mr. Willis, however, treats the 'idea' as implying the 'complete homogeneity of nature'; but that surely is not a postulate of methodology, but a metaphysical doctrine. If so, it could not be 'demonstrated'; nor have we any reason to suppose it true[3].

[1] The demonstrations are almost pedantically set out in the form of numbered questions, subquestions, postulates, lemmata, theorems, and corollaries. The alternative hypotheses—that the *maculae* and *faculae* may be due to opacities in the eye, in the lenses, in the atmosphere, or to planets passing between the earth and sun—are considered in turn and formally disproved; and the one remaining hypothesis—disturbances of the sun's own surface, which rotate with the sun itself—is systematically confirmed. Scheiner calculates the size of the spots in square miles. From the rate of their passage across the disc he correctly computes the rate of the sun's rotation (approximately 27 days as against Galileo's cruder estimate of 30 days); from the fact that their paths vary progressively from month to month (rectilinear but oblique in December, concave northwards in March, etc.) he correctly infers an annual shift in the position-angle of the sun's axis; and from the fact that their angular speed is greater near the equator than in higher latitudes, and that they appear to be cooler and therefore darker depressions which change their shape, he argues that the sun is not a solid body (as Galileo assumes), but a rotating sphere of flaming gas (*ex nube ignea et lucida, liquida flamina*). The whole argument is prefaced by the relevant preliminaries—an exposition of the optic principles underlying the new 'optic tubes' (based, he tells us, on the work of Roger Bacon, 'vir Anglus non tantum in mathematicis, sed in omnibus aliis scientiis eminentissimus') and a systematic discussion of the theory of stereographic projection (since the sun appears flat but is actually spherical). In short, as his German biographer observes, Scheiner works his way to his final conclusions by dint of *echt deutsche Fleiss*, whereas Galileo is guided largely by his brilliant *divinatorische Kraft*. The engravings of scientists actually at work will alone reward the student who takes the trouble to hunt up this neglected but fascinating volume.

[2] At heart (as Father Kircher told Gassendi), he was a Copernican. As a Jesuit, however, Scheiner was bound in his published works to disown the heliocentric doctrine. But, except on this issue, which he rather shirks, he seems to come off best in his prolonged controversies with Galileo.

[3] Mr. Willis, for example, tells us that the principle of homogeneity requires us to assume that "the sun and stars . . . consist of matter of precisely the same kind in precisely the same state as the matter composing the objects familiar to us on the earth". Would he say that this must hold good of matter at the centre of the sun or in one of the White Dwarfs—the companion of Sirius, say, or Kuiper's star? The latter has a density several million times that of matter in its solid phase on earth, and yet possesses the properties of a gas: a thimbleful would weigh several thousands of millions of tons, and its temperature is about 18 million degrees, so that the gas—a 'degenerate'

The fact is that, as Heisenberg and other physicists have argued, the ideal of a single unified science, forcing all observable phenomena into one vast homogeneous system—the mechanical universe of the nineteenth century materialists—has proved quite untenable. "Science," as he says, "no longer claims to make assertions about nature as a whole": instead it is content to adopt "a more or less piecemeal approach, in which certain broad groups of phenomena are isolated and treated within an admittedly limited framework."[1]

In spite therefore of all the arguments of Mr. Willis and Lewin, we feel that, so far as methodology is concerned, the modern psychologist has little to learn from the writings of Galileo. Galileo's outstanding contributions to science lay in his development of such novel concepts as 'inertia', 'uniform acceleration', and 'velocity at a given moment' in a body whose velocity is continually changing: but these are instances of original insight, not of a new methodology or technique. Were we asked to name from these early days a model of scientific method, it would not be one of Galileo's dialogues, but a research by one of his greatest contemporaries at the University of Padua, Harvey's *De Motu Cordis et Sanguinis*. Here Mr. Willis would find all the devices on which he insists—planned experiment, specially constructed instruments, and the systematic use of measurement and calculations based thereon; moreover, the whole argument is a model of the inductive procedure, based on the so-called 'hypothetico-disjunctive' proof, with the rival hypotheses (Galen's and his own) formally stated, and the one disproved and the other confirmed.

Scientific Laws versus Statistical Laws. Dr. Taylor, like Mr. Willis, quotes with approval Lewin's pronouncements on the "conflict between Galileian and Aristotelian modes of thought in contemporary psychology", and concludes that "what psychology needs is not 'statistical laws' but 'scientific laws'"[2]. He does not explain what he understands by these terms, or why they should be so sharply opposed. The antithesis, I imagine, is based on a distinction drawn by Eddington, though it is to be found even in Aristotle's own teaching. 'Laws of nature', says Eddington, may be of two kinds; (i) first, 'deterministic laws', which make statements that are (in Aristotle's phraseology) 'exact', 'certain', and 'eternally' true; (ii) secondly, 'statistical laws' that make statements that are merely true ἐπὶ τὸ πολύ, i.e. they state only what is probably true. "So long as the aim of the physicist was to bring to light a deterministic scheme,

gas—must be almost completely ionized. Matter in such a state can only be discussed in mathematical terms; and Mr. Willis's argument would lead us to reject the whole notion as contradictory to his fundamental postulates. The corollary is obvious. Matter which is conscious (if there is such a thing) may also be quite unlike matter (or the concept of matter) familiar to us from the studies of the older physicists; and certainly the psychologist need no longer feel bound, as Lewin supposed, to treat man himself as just another material body with no other properties except those found in material bodies that are unconscious and inanimate.

[1] W. Heisenberg, *The Physicist's Conception of Nature* (1958, pp. 180f.): Heisenberg incidentally points out that one of the distinctive features of Galileo's own work was his rejection of the cramping limitations imposed by the traditional view of the universe as a single system, and dealt instead with "modestly delimited sectors extracted from the multiplicity of phenomena".

[2] 'Experimental Design; A Cloak for Intellectual Sterility.' *Brit. J. Psychol.*, XLIX (1958) pp. 106.

the pursuit of these 'secondary' laws'' (says Eddington) ''appeared a blind alley, since they lead only to probabilities: on the modern theory both kinds of law belong really to the statistical type.[1]'' It is this 'modern theory' that Dr. Taylor wishes to reject. In particular, he maintains, it is wholly fallacious to conclude, as recent writers have done, that ''psychology is an *inexact science*'' (his italics).

As an ideal 'law of nature', with nothing statistical about it, Dr. Taylor takes Newton's equation for determining the 'force' of gravity ($Gm_1 m_2/r^2$). Burt [2] cited the same equation; but, as Mr. Willis points out, while Burt treats it merely as a provisional hypothesis, Dr. Taylor regards it as a final law, 'exact and universal': from it, he says, we can predict the motions of the planets with a success which mere statistical inferences from ''facts about each of a thousand solar systems collected by a thousand observatories'' could never have achieved.

To this interpretation there are obvious objections. First, if we suppose (with the majority of contemporary physicists) that the 'force of gravity' is itself merely a macroscopic manifestation of some more fundamental process (e.g. the exchange of mesons between the two bodies), the effect becomes by its very nature a statistical resultant. In any case, although G is called a 'universal' constant, its value can only be estimated empirically by repeated observations, not perhaps of ''a thousand solar systems,'' but of many hundred physical systems, terrestrial as well as celestial; and the fact that it is always cited with a probable error attached indicates that its investigators do not share Dr. Taylor's objections to statistical procedures. Moreover, there would be no cogent reason for assuming that Newton's law holds good of other systems besides our own, unless data were available from other systems (e.g. the remoter binary stars). And, as it turns out, where velocities are high, the law does *not* hold good.

Finally, even with the most accurate value available for G, Newton's equation does not allow us to ''predict the motions'' of the planet Mercury: the discrepancy is over 30 times its probable error. Leverrier, who had already deduced the existence of Neptune from the irregularities of Uranus, was consequently led to deduce from Newton's law the existence of yet another planet, Vulcan. So far, Vulcan has eluded all discovery[3]. And Einstein's new law of gravitation

[1] *New Pathways in Science*, p. 80. Cf. Aristotle, *Phys.*, II, 4–6. It is interesting to note that Aristotle himself inquires whether 'chance' (i.e. 'accident' or 'sheer contingency') should also be regarded as a veritable 'cause' (in addition to the four main types of causation that he has already recognized). In the end he concludes that it should not: every event follows from determinate causes of its own, and chance is simply a name for ''the unforeseen meeting of two or more rigorous chains of causation''. Thus, as Ross observes, ''we have no reason to attribute indeterminism to Aristotle''. And certainly in the days of Galileo the objections to the use of probabilistic arguments would have come from the Aristotelians not from Galileo himself, who actually contributed to the theory of probability (*Opere*, III, pp. 119–121, XIV, pp. 231–284). On this point, therefore, Lewin and Dr. Taylor are themselves following Aristotle rather than Galileo.

[2] 'Definition and Scientific Method in Psychology,' this *Journal*, XI (1958), pp. 46.

[3] Dr. Taylor replies that the non-existence of Vulcan is a non-statistical fact. But one can never be *absolutely* sure that the non-existence of a thing *is* an actual 'fact': it is at best only a probability. A non-reflecting body might easily baffle all observers. The reason why astronomers have given up the search is that the deductions from Einstein's law fit the data far more closely than the deductions from the hypothesis of an additional planet.

successfully predicts the unusual movements both of Mercury and of bodies with exceptionally high velocities.

Clearly then it is no longer possible to think of Newton's law as an exact or universal law of nature; it is merely a 'law' (or rather a postulate) describing a particular model, which, for the purpose of certain *approximate* predictions, we find it convenient to accept. There seems nothing in Dr. Taylor's arguments, therefore, which disproves the view that *some* 'scientific laws' at least are in point of fact 'statistical laws', or which compels us to reject 'statistical laws' as essentially unscientific[1]. Accordingly, the psychologist remains free to adopt such a law when the balance of evidence favours it.

II. The Hypothetico-Deductive Method

The Fallacy Involved. Mr. Willis, like other loyal Lewinians, acclaims the hypothetico-deductive method as the "sovereign rule for science", and ascribes its origin to Galileo. The method, he says, "satisfied scientists from Galileo to Russell . . . . Where then does the fallacy lie?" The fallacy is that which Aristotle long ago pinned down as the 'fallacy of the consequent' (τὸ [σόφισμα] παρὰ τὸ ἑπόμενον, *Soph. Elench.* vi, 168b 27, cf. xxviii, 181a 27). It arises whenever confirmation is mistaken for proof. Galileo himself supplies an example: "If the speed of falling bodies is proportional to their weight, a heavy body would reach the ground sooner than a light. Observation shows that this is so. Yet this does not prove the premiss, for the lighter body may be retarded by the resistance of the air". The error is exposed in every textbook of logic; and consists in assuming $a \supset b \equiv b \supset a$: or, as Joseph put it, "in treating a condition and its consequent as convertible". And, as he goes on to point out, it is one of the commonest causes of invalid reasoning and of the acceptance of scientific theories that are entirely erroneous (*Introduction to Logic*, pp. 596–7).

Neither Galileo nor Newton relied on merely verifying their hypotheses. When we turn to their actual writings, we find that almost invariably they also sought to disprove the chief rival hypotheses[2]. Today these discredited

[1] Even before the advent of relativity and quantum theory, however, modern physics had to accept some laws which were purely statistical, e.g. the second law of thermodynamics. Would Dr. Taylor reject this 'law' as wholly 'unscientific'?

[2] It was in point of fact the earlier English empiricists who directed attention to the "method of elimination", with their recurrent formula: "sed hoc est falsum, ut patet experimento". As Crombie observes, "Grosseteste's great contribution was to emphasize the importance of falsification, and to develop it into a systematic method of experimental procedure" (*Robert Grosseteste and the Origins of Experimental Science*, p. 84; see the whole passage). Francis Bacon adopts it as the keynote of his method. "Maior est vis instantiae negativae" (*Novum Organon*, II, aph. xvi): and again (in the passage quoted in *The Factors of the Mind*, pp. 34f.) "the first task of induction is rejection or exclusion".

Apparently on the authority of Mr. Jameson, Mr. Willis asserts that, in *The Factors of the Mind*, Burt "advocates the hypothetico-deductive method in psychology". But, if he will look at the chapter cited, he will see that the method advocated was not 'hypothetico-deductive' (in his sense of the phrase), but 'hypothetico-disjunctive' (p. 37). The difference will be clear from the examples quoted of actual researches. It is unfortunate that so many writers confuse the two procedures, and often, it would seem, use the former technical name when they really mean the latter. Not every 'method' that involves a hypothesis and a deduction is to be labelled a 'hypothetico-deductive method' any more than every method which involves physics and psychology would be called a 'psychophysical method'.

alternatives are no longer of interest. Hence the modern textbook fails to mention them, and so gives a somewhat misleading impression of the gist of the original proofs. They were nearly always of the type we have already referred to as 'hypothetico-disjunctive' (cf. *loc. cit. sup.*, this *Journal*, XI, p. 66).

An Illustrative Example. By way of illustrating how his methodological doctrines apply to current psychology, Mr. Willis cites a research recently published by Dr. Hildebrand. This, he says, affords a "typical instance of the success with which the hypothetico-deductive method may be employed". Although we readily acknowledge the skill with which Dr. Hildebrand has approached his difficult task, we venture to suggest that his inquiry really provides a typical instance of the weaknesses of the method he has chosen. It is a method that seems particularly popular with factorists, especially those who insist on 'rotating axes to meaningful positions'. Yet it is precisely in this field of work that its use is likely to be most misleading.

The preliminary inferences that Dr. Hildebrand draws from his own hypothesis are perfectly valid, namely, that, if the proposed classification of neurotics is sound, then the various groups that he has tested could all be reduced to three categories only (neurotic extraverts, neurotic introverts, and normals); and that, in such a case, the average factor-measurements calculated for the several categories would reveal differences that are statistically significant. After analysing the results obtained, he tells us that "all the deductions have been supported" (with one minor exception); the tests therefore are "successful", and the theory proposed "confirmed". But here the argument depends on inverting the intitial deduction; and to do so is surely to reason like the doctor in Molière's play: "If the gentleman has the fever, he would be crimson in the face; but he *is* crimson in the face; therefore he has the fever". In both cases, it is plain, there may be several other hypotheses that would account for the phenomena observed quite as plausibly—perhaps even more plausibly—than the one suggested. Dr. Hildebrand makes no attempt to disprove the alternative classifications that have been put forward, and ignores the criticisms urged by rival theorists against the particular scheme that he himself favours[1].

The name goes back to a classification of scientific methods introduced by Mill and his followers: viz., I. Inductive, II. Deductive, A. Abstract (as in geometry), B. Concrete, i. Direct, or Hypothetico-deductive. ii. Indirect or Empirico-deductive. Mill, it should be remembered, held that strictly inductive methods were impracticable in the social sciences. Jevons, however, maintained that, rightly understood, induction was always hypothetico-deductive (Mill, *Logic*, I, bk. iii, chaps, 1, 11, 14, and II, bk. vi; Jevons, *Principles of Science*, pp. 265f.). In actual practice the hypothetico-deductive procedure commonly owes its plausibility to the fact that far more is known or taken for granted than is explicitly stated in the inference as formulated.

[1] Criticisms both of the hypothesis and of the methods adopted will be found in the review that Mr. Willis cites (this *Journal*, V, pp. 208–212). He goes on to ask: "What other alternative can he (Gregory) suggest in its place?" The most obvious (mentioned by Gregory in a later note, *ibid.*, VI, p. 51 and refs.) is the scheme put forward by Burt in 1916, namely, the hypothesis of a *general* factor of 'emotionality', supplemented by a *bipolar* factor contrasting inhibitive or repressive tendencies with uninhibitive or unrepressed. The *factor* which Dr. Hildebrand designates 'neuroticism' shows some resemblance to the factor of 'general emotionality'; but the *condition* designated 'neuroticism' seems to imply a state that is essentially pathological. Burt's factors are

The underlying assumption seems to be that, if a test or a rotated factor shows a positive and significant correlation with the hypothetical condition it is intended to measure, then this is sufficient to ' confirm ' the hypothesis in question. But of course the crucial point—particularly in a research with practical applications in view—is not the mere presence of a correlation or a factor loading, but its size; and it must be admitted that the size of the figures in Dr. Hildebrand's tables is decidedly small. Moreover, when axes are rotated in order to "adjust them to a meaningful position", it is all too easy to secure a fairly plausible correspondence with the factors postulated by the hypothesis to be confirmed. By the ordinary summational method Dr. Hildebrand first obtains *three* centroid factors, which he finds fully significant when tested by Burt's formula. He then rotates them, and discovers that *four* will be needed to achieve the requisite agreement. Two of the rotated factors (which have decidedly low factor variances—only 12 per cent and 6 per cent respectively) are identified with the two required by his hypothesis, namely, neuroticism and intro-extraversion. But to effect this two supplementary factors have to be introduced; the first is identified with ' general intelligence ', while the other baffles all attempts to give it a meaning. Plainly, however, if one is allowed to introduce meaningless factors to account for the residuals, any set of correlations can be fitted with almost any pair of ' meaningful ' factors we like to choose[1].

Yet even now the actual verification is far from complete. For one thing, as Dr. Hildebrand admits, "the normal group did not in fact fulfil the prediction made for it": the normals appear more extraverted than the ' hysterics ' or ' psychopaths ', whose condition (according to the hypothesis) is the consequence of excessive extraversion. Most psychoanalysts would expect the obsessionals

common to both neurotics and normals. But with genuine neurotics there is in addition a ' pathological group factor ', which implies a qualitative (and not merely quantitative) difference between them and the normal and a series of smaller and more specialized factors furnishing the differential diagnoses: (cf. this *Journal*, V, p. 210). We agree with Jung that the mere ' conflict ' of emotional forces, with one set tending to oppose the other, does not in itself constitute neuroticism, as the clinician understands the term; nor would the condition of ' emotional tension ', which results from a high degree of ' emotional energy ', suffice of itself to produce a neurotic state. There is, as it were, a ' critical point ' at which a fresh state supervenes: equilibrium breaks down, and a new and abnormal condition of ' instability ' ensues. The model we have suggested is that of nerve currents passing from a ' steady state ' of ' stream-lined ' flow to a state of ' turbulence ' (the terms being used with the technical meanings they bear in hydrodynamics). To test such a hypothesis, however, something more than the merely static analysis of ordinary factorial procedures would be required.

There is, however, yet another ' hypothesis ' that at least deserves consideration, namely, that the tests employed by Dr. Hildebrand had very little value as methods of clinical diagnosis: few psychiatrists would place much reliance on a patient's performances in a ' lie-scale ', a test of ' sex humour ', ' lifting pencils ', ' body-index ', ' progressive matrices ', or the like; and certainly the low correlations and factor variances seem to support this alternative conclusion.

[1] Incidentally we should like to deprecate the custom of printing rotated factors only. From the figures given by Dr. Hildebrand we have ourselves endeavoured to reconstruct the observed correlations, and have then refactorized them in the usual way: we obtain two main factors—a general and a bipolar; these might well be identified with ' general emotionality ' and ' inhibition *versus* its opposite '. And, since their factor variances are much larger than those given by Dr. Hildebrand for his rotated factors, we might argue that our alternative hypothesis receives a much stronger confirmation than the hypothesis that he himself prefers.

also to be of an extraverted type: yet, according to the factorial results, they are more introverted than the 'anxiety' cases. Finally, Dr. Hildebrand himself mentions two disappointing features in the evidence: "first, although the experiment was designed to amplify the I–E factor as much as possible, its total variance amounts to only 6 per cent—only 1 per cent more than the factor it was found impossible to identify: secondly, of the tests suggested as possible measures, only two differentiated between introverts and extraverts at the 1 per cent level of significance". Indeed, if we may judge from the variances given, nearly 40 per cent of the cases would have been misclassified. That being so, it is a little surprising to find him in the end describing his inquiry as 'successful' and 'satisfactory[1]'.

The Formulae for the Probability of a Hypothesis. Mr. Willis next inquires "how the formula [quoted by Mr. Gregory for the probabilities] emerges from his theory of inductive reasoning". The formula was intended simply as a quantification of the hypothetico-disjunctive argument. The underlying principle was fully discussed and defended in *The Factors of the Mind* (pp. 37f.). Taking the simplest type of problem, the algebraic proof may be outlined as follows.

Let P_h denote the final or posterior probability of the hypothesis[2] to be maintained, r the result of the observations, p the antecedent probabilities (based on the information available before the result was obtained), q the likelihoods (i.e. the probabilities of the result if this or that hypothesis were true): then, assuming that there are only two hypotheses, namely, h and its negative n,

$$P_h = \frac{p_h \cdot q_h}{p_r} = \frac{p_h \cdot q_h}{\sum p_i \cdot q_i} = \frac{p_h q_h}{p_h q_h + p_n q_n}. \qquad \text{(i)}$$

(p_r is equated to $\sum p_i q_i$, because the alternative hypotheses considered are assumed to be mutually exclusive and jointly exhaustive.) Dividing throughout by $p_h q_h$, we obtain

$$P_h = \frac{1}{1 + p_n q_n / p_h q_h}, \qquad \text{(ii)}$$

which is the formula cited by Gregory (this *Journal*, VI, p. 109).

[1] We thought it only right to submit both Mr. Willis's comments and our own to Dr. Hildebrand. He writes that he himself did not mean to claim that the hypothesis proposed was proved: "I merely stated that the *deductions* which I had arrived at—perfectly correct according to Gregory—had been proved, and that additional support had thus been given to the hypothesis itself". But, unless there is some independent evidence that "tests *purporting* to measure I–E and neuroticism" (our italics) really *do* measure these characteristics, Dr. Hildebrand's deductions are not really 'proved'. In the past the usual method for verifying this particular classification by means of factorial analysis has consisted in factorizing correlations between direct observational ratings for the behavioural characteristics included in the definitions of 'neuroticism' and 'I–E' (intro- and extraversion). At best an analysis of test-results could only support the hypothesis that the tests Dr. Hildebrand has used really do measure the traits he hoped they would measure: it could hardly yield 'additional support' to conclusions derived from more direct assessments.

[2] As commonly used the phrase 'probability of a hypothesis' may mean either (i) the probability of a hypothetical proposition (as here) or (ii) the probability of a hypothetical cause. To derive the latter from the former a further step is essential which is too often slurred over under cover of the ambiguity which thus arises.

From this it will be clear that the probability in favour of the hypothesis will be greatest when (i) p_h and q_h are as large as possible and (ii) p_n and q_n as small as possible. Maximizing p_h means that, in planning his experiment, the investigator should choose first the hypothesis which has the greatest antecedent probability or (if the essential choice is already fixed) should see that it is expressed in such a form that this condition is ensured. When he has thus maximized p_h, he has automatically minimized p_n (since eqn. (ii) assumes only two alternatives). Maximizing q_h means that he must also choose for his experimental test the corollary to which his hypothesis allots the highest likelihood; in some cases it may be possible to select a corollary which is directly deducible from his hypothesis, i.e. which makes $q_h = 1$. To maximize q_h, however, is *not* equivalent to minimizing q_n. Hence his third aim should be to choose for verification an observable result which would be highly improbable were his hypothesis false.

In the vast majority of researches, however, little or nothing will be known about the quantitative values of the antecedent probabilities, and the likelihoods will have to be estimated from the data. In that case the traditional procedure is to assume that the antecedent probabilities are equal; and this is tantamount to adopting the principle of 'maximum likelihood'.

The Consilience of Inductions. Those who advocate the hypothetico-deductive method commonly speak as though just one experiment, which successfully confirmed the hypothesis, would provide a conclusive test. That would seldom be deemed sufficient in other branches of science; and in a complex science like psychology it would be decidedly precarious[1]. Wherever practicable, the investigator should aim at securing as many confirmatory results as possible, each independent of the rest. Complete independence may be out of the question; but, even if it is only approximate, its cumulative effect is surprisingly high. The theoretical formula is as follows.

If N independent corollaries have been independently verified, and $P_1, \ldots P_N$ are the several posterior probabilities, and p_r the antecedent probability of the conclusion on which all agree, then the cumulative probability (P_c) resulting from this 'consilience'[2] will be

$$P_c = \frac{P_1 P_2 \ldots P_N}{P_1 P_2 \ldots P_N + (1-P_1)(1-P_2)\ldots(1-P_N)(p_r/[1-p_r])^N}.$$

In practice the conditions assumed in proving the foregoing equations will rarely be fulfilled with any exactitude, and the values to be inserted may often be unknown. Nevertheless, though actual calculation may be impossible, the formulae still yield a vivid illustration of the ways in which the credibility of an 'inductive' conclusion can be augmented by suitable planning.

[1] See Burt, this *Journal*, XI, pp. 42, 61f.

[2] The phrase 'consilience of inductions', which seems to have puzzled Mr. Willis so much, was borrowed from Whewell, *Novum Organon Renovatum*, pp. 95f.

III. The Demand For Causal Laws

The Implications of Quantum Theory. Mr. Willis's third contention is that, since "science always reveals strict causal determination, psychology, if it is to rank as a science, must also aim at discovering exact causal laws". His reasons seem to be based, not so much on logical or scientific grounds as on metaphysical. In effect he appears, like Dr. Taylor, to be still clinging to Kant's unfortunate doctrine that causality is an *a priori* category.

He begins his argument by criticizing "the probabilistic hypotheses advanced by the later quantum theorists", and the inferences that Burt drew from them in a recent paper in this *Journal* (XI, pp. 77f.). His own attitude towards these 'revolutionary views' might be summed up in the words used by Einstein in an early letter to Bohr: "Du glaubst an den würfelnden Gott, und ich an volle Gesetzlichkeit: der grosse anfängliche Erfolg der Quantentheorie kann mich doch nicht zum Glauben an das fundamentale Würfelspiel bringen". Mr. Willis argues, as Einstein did, that what to us seems an irreducible indeterminacy is really an effect of our own human ignorance: underlying the observable manifestations there *must* be certain 'hidden causal variables', presumably at a sub-quantum level, which have hitherto defied discovery. But such an interpretation, as has so often been pointed out, is wholly incompatible with the quantum theory in its generally accepted form and quite inconsistent with the experimental results[1].

However, indeterminacy is by no means the only paradoxical feature that seems to characterize both the world of conscious experience and the physical world as the quantum theorist sees it. In the paper that Mr. Willis criticizes it was singled out merely because it was directly relevant to the issues raised by Mr. Clarke and Mr. Jackson. Mr. Willis goes on to ask "what precisely are the analogies which Burt discovers, or thinks he can discover, between quantum phenomena and mental phenomena?" There are, we believe, four crucial characteristics which differentiate the conceptual models adopted in quantum theory from the models adopted by the classical theory and which are closely parallel to those suggested by the results of psychological research.

1. In the classical theory it is assumed that processes going on in the universe may be validly analysed in terms of separate parts, each with its own intrinsic and independent properties. In quantum theory processes are treated as indivisible wholes (and indeed as aspects of one great indivisible whole); and the distinguishable components appear merely as conceptual abstractions. Thus quantum theory, quite as much as 'Galileian physics' (if we may borrow Lewin's phrase), insists on the 'unity of the universe'. But it does not for a moment suppose that a unitary universe must of necessity be homogeneous: qualitative differences are regarded as irreducible, and no less important than quantitative. Moreover, the unity which it postulates is not a mechanical but rather an organic

[1] For the more important instances, see D. Bohm, *Quantum Theory*, pp. 611, 622f. Einstein's argument is stated in *Phys. Rev.*, XLVII, 1935, pp. 777f.; for the replies by Bohr and others, see *ibid.*, XLVIII, pp. 696f., XLIX, pp. 393f., 476f.

unity; the universe therefore is not a mere additive whole which can be dissected into isolated parts, but a whole within which we may recognize distinguishable aspects. To deal with one aspect at a time is a convenient rule of practicable methodology: but such treatment necessarily involves abstraction.

All this is closely analogous to the standpoint of modern psychology. As the Gestalt school has taught us, conscious experience cannot be cut up into separable items or elements, each independent of the other, as the old 'brick-and-mortar' school assumed. It forms an indissoluble whole, and an organic whole, not a mere sum of parts—a whole in which distinguishable yet inseparable aspects and phases can be discriminated and singled out for special attention. No doubt, the behaviourist will reply that the characteristics of a man's private consciousness are not a subject for scientific study, since science is concerned solely with "the structural and behavioural characteristics of material systems". Thus, while the behaviourist wishes to abandon one of the most important sources of psychological evidence in order to work with the over-simplified models of the classical school of physics in which he was trained, the physicist of today proposes to abandon the seductive simplicity of the older theories in order to work with what is essentially a theory of Gestalten.

2. The quantum theorist's insistence on the 'unity of the universe' (as he interprets the phrase) brings with it a corollary of outstanding importance. In analysing the results of the data furnished by his observations, the modern physicist finds that he can no longer ignore the observer. The classical theory treated any given system that happened to be the object of investigation as something which could be completely isolated from its environment, including the observer and his measuring instruments. Quantum physics treats system and environment as indissolubly linked: the mere effort to observe and measure a particle is held to interfere with that particle's behaviour.

Precisely the same situation arises in psychology. "The very act of observing our own conscious states," says McDougall, "tends to modify whatever experience we try to examine and describe: the smaller, the fainter, the more elementary the process under examination, the greater, relatively speaking, is the disturbance and uncertainty produced by focusing the full light of introspection on to it." The behaviour of a child at an examination or an interview, the behaviour of an adult in the artificial surroundings of a laboratory, and even the observable experiences of those psychologists who have been thoroughly trained in the difficult technique of introspection, all differ in varying degrees from their behaviour and their experiences in more normal circumstances. And, as most psychologists would agree, this is not something that can eventually be eliminated by further training or improved techniques; it is, as Ward puts it, "inherent in the very nature of the relation that holds between object and subject—between observed and observer—in the unitary process of awareness".

3. The classical theory assumed that all observable phenomena could be expressed in terms of discrete and separable particles, and that all modifications in their states would exhibit a perfect continuity: the differences, as Lewin was so

fond of reminding us, were differences solely in quantity, not of quality. Thus, in describing the motions of a particle, the classical theory treated the two fundamental variables—position and momentum (or velocity)—as capable only of continuous variation. Quantum theory has been forced to reverse all this. An intelligible meaning (so it is maintained) can only be given to momentum, when we postulate, not a particle sharply defined like a geometrical point, but a particle thought of as covering a small but indefinite range of positions, while its average position changes in a way that can best be represented by the model of a wave. Moreover the transitions from one quantum ' state ' to another are abrupt and discontinuous: in popular phraseology, the electron ' jumps ' from one orbit to another. Hence, as Eddington puts it, " the continuity assumed by the traditional theory proves to be an anthropomorphic illusion arising when we are confronted with macroscopic—i.e. complex, man-sized—phenomena ".

These paradoxes are again analogous to those that have been noted by almost every modern psychologist who does not persist in ignoring the results of introspection. Impressed by the apparent success of the particulate and atomic theories in nineteenth century physics and chemistry, the earlier British psychologists attempted a similar analysis of all ' conscious contents ' into certain basic ' elements ' called ' ideas ' or ' sensations '. Behaviouristic psychology substituted ' reflexes ' or ' responses to stimuli ': but in either case the standpoint was analytic. This however distorts the obvious facts. As the later British School (Ward, Stout, and McDougall) were never tired of insisting, our ' sensations ' are not separable particles of consciousness, but mere abstractions with blurred edges and an indefinite duration.

At the same time, beneath the apparent continuity of conscious experience, we discover unexpected discontinuities. At the cinema, in experiments on the phi-phenomenon, and in many other equally familiar cases, we have a clear experience of continuous motion, yet the stimuli are discontinuous[1]: though the stream of consciousness seems continuous, the underlying neural processes are now known to consist of discontinuous, all-or-nothing discharges. Moreover in their qualitative nature the states of consciousness themselves present certain curious discontinuities; no smooth transition can possibly be conceived which would enable us to pass by gradual changes from an auditory sensation to a visual: you lie in bed with your eyes closed, and are aware of noises, odours, and the like; you open your eyes, and an entirely new state of consciousness abruptly appears—

[1] On closer study it is really the assumptions of the classical theory that turn out to be paradoxical. It is in point of fact almost impossible to secure a distinct conception of continuous change as such. Based as it was on the differential calculus, the notion of continuous change assumed that we could intelligibly think of a velocity at an *instant* of time or a *point* in space, i.e. that $\Delta x/\Delta t$ approaches a definite limit as Δt is made indefinitely smaller. The paradox is as old as that of Zeno's arrow, which " cannot move where it is, and cannot move where it is not ". It would be far easier to assume that, when Δt becomes smaller than a certain specific amount, the ratio *ceases* to tend towards any limiting value.

a world of light and colour where the qualities are entirely disparate from those supplied by the other senses[1].

4. In physics the classical theory treated the motions of a particle as completely determined by the equations of motion; and, when psychology changed from a branch of philosophy into a branch of natural science, practically all its earlier exponents assumed that mental changes must likewise be subject at bottom to a rigorous causal determination. Yet it was quickly discovered that this mode of treatment would bring with it two great difficulties. First, the hypothesis of complete mechanical determination is in flagrant contradiction with the direct experience that we have (whether illusory or not) of being able ourselves to control events by rational or non-rational choice: this is a point which we have sufficiently emphasized in previous papers, and need not be further elaborated here. But secondly the deterministic hypothesis of the physicist makes nonsense of our experience of time. In popular discussions physical determination is regarded as essentially causal, resulting from the operation of forces; and causation (in this sense) implies that time has a one-way direction. But in classical physics the determination works backward as well as forward: (this is obvious from the specimen laws that Dr. Taylor cites—the first law of motion and the law of gravitation). Hence cause, as the modern physicist recognizes, becomes inappropriate or at least irrelevant.

The dimension of time, as we experience it, is quite different from the dimensions of space. In space, direction is, from a theoretical standpoint, indifferent: we can, as the factorist says, ' reverse the signs ', and the results remain the same. But events in time are irreversible; and to introspection at any rate this characteristic seems to defy elimination both at the perceptual and at the conceptual level. Mirrored in the looking glass, the world of space still makes sense; but mirrored time, as is so vividly shown by Lewis Carroll in *Silvie and Bruno* and Wells in *The Time-Machine*, would turn the processes of the world we know into a meaningless farce. And the modern substitution of a four-dimensional space-time merely enhances the anomaly.

In classical physics, however, a one-way dimension of time makes no appearance until we reach the second law of thermodynamics (stating increase of randomness); and this is essentially a statistical law. Since Dr. Taylor rejects all statistical laws as wholly unscientific, it is by no means easy to see how he would cope with the psychology of time. In quantum theory, since there are no deterministic laws, but only laws stating probabilities, the difficulty no longer arises; and we can still talk (if we wish) in terms of causation, with momentum substituted for force. But in that case we should have to regard the causes as expressing general tendencies or trends, not the entire behaviour of the system. And this, of course, is equally true in practical life. There, when considering the conscious

[1] It is curious to recall that, while Lewin was claiming that the Galileian methodology in physics had wholly replaced qualitative variations by quantitative, Eddington was defending Leibniz's view that " the science of quantity, dealing solely with numerical relations, should be subordinate to the science of quality, which treats of similarity and order ". And in this respect he has been followed by many later physicists.

behaviour of any given individual, we find the notion of complete determination appearing only as a limit. We can recognize tendencies, impulses, and trends; but we always allow for the possibility that unforeseen and apparently uncorrelated factors may prevent such tendencies reaching their natural goals[1].

The foregoing considerations will help to indicate the kind of model that will be most appropriate in quantitative psychology, and the type of mathematical equation to be used in describing it. The choice lies between two general methods of analysis. The enthusiasts for a Galileian methodology apparently assume the validity and the superiority of what they would doubtless call ' Galileian ' analysis, but which should perhaps rather be termed ' Newtonian '. We ourselves would suggest that the more appropriate analysis is that which is common to the theory of sets, the theory of relativity, and the theory of quantum mechanics: the most convenient name is that generally used by continental writers—' Hilbertian ' analysis.

Dr. Taylor, we have seen, suggests, as a typical equation for the psychologist to imitate, Newton's inverse square formula for his ' law of universal gravitation ', viz.,

$$F = G \cdot M_1 M_1 / s^2,$$

where M_1 and M_2 are the masses of the bodies concerned, s the distance separating them in ordinary three-dimensional space, and F the force produced (all being measured in some consistent system of units, e.g. dynes, grams, and centimetres), while G is a ' universal constant ' depending on the units employed (e.g. $6{\cdot}67 \times 10^{-8}$ in the c.g.s. system). This and the related equations of Newtonian dynamics assume absolute three-dimensional space and absolute time; and the characteristic mathematical tool is the differential calculus, or (as Newton called it) the ' method of fluxions '. Unless the quantities concerned can be measured in absolute units any such equation is meaningless. In the formula above each of the three variables can be measured on a ' ratio scale ', i.e. a scale in which it is possible to determine not merely (i) order (i.e. greater or less), but also (ii) equality of differences, and (iii) equality of ratios[2]. Now, in modern psychology it is only in

[1] We suspect that at bottom Dr. Taylor's reasons for the views he so vigorously champions are really metaphysical. In effect he confronts us with two models of the universe: first a universe, which is wholly governed by universal laws of nature, and in which therefore, to an omniscient eye like that of Laplace's perfect mathematician, all events would be seen as logically determined by the initial conditions—a universe in which ' randomness ' therefore would be merely a name for human ignorance; and secondly, a universe which includes contingency as well as rational determination and in which the future is *really* indeterminate, so that, even in principle, ignorance could never be reduced below a certain limit. The former—the model of classical physics—has a strong appeal for the technologist; the latter for the more diffident mind of the research-worker. After all, classical physics arose primarily out of the attempt to provide a rational basis for technology—steering ocean-going vessels, aiming projectiles, manufacturing clocks, etc. To this day it is still concerned largely with machinery; and what designers and manufacturers hope to produce are machines whose operations shall be completely and precisely determined, so that all error is abolished. Their ideals therefore demand a thoroughgoing determinism as a basic postulate. The ideals of the research-worker are quite different; and, since human beings do not behave like man-made machines, the scientist who undertakes research in psychology is in the end almost bound to adopt the second and humbler type of model.

[2] Cf. C. Burt and C. Banks on ' The Nature of Mental Measurement ', ap. C. A. Mace and P. E. Vernon (ed.), *Current Trends in British Psychology* (1953), pp. 158f.

exceptional situations that measurements of this kind are possible. Often we can do little more than arrange the objects or their properties in ascending or descending order. Evidently if that were the most we could do in the case of mass and distance, we could write our equation in the simpler form $F = G \,.\, M_1 M_2/s$.

The difficulty is still more obvious if we take the equation which Galileo himself claimed to deduce from his experiments with balls rolling down an inclined plane. His final hypothesis, it may be remembered, was that the distance traversed varied with the *square* of the time: i.e. in symbolic form,

$$s = \tfrac{1}{2}at^2,$$

where s as before denotes distance, t time, and a the acceleration due to gravity. The course of the theoretical trajectory, plotted as a continuous curve, will be expressed analytically by determining the values for the successive derivatives. We then obtain

$$\frac{ds}{dt} = at \text{ (velocity)}, \qquad \frac{d^2s}{dt^2} = a \text{ (acceleration)}, \text{ and } \frac{d^3s}{dt^3} = 0:$$

(cf. this *Journal*, XI, pp. 40f.). If, however, the variables were psychological measurements, such a treatment would rarely be possible. Suppose, for instance, that, instead of representing the successive positions of a ball moving in space, s represented the positions of pupils assessed for their progress in arithmetic, and t their rank-order for (say) industry, and that we then wanted to express the relation between the two in a Galileian formula: there would be no means of discriminating between $s = \frac{1}{2}at$ and $s = \frac{1}{2}at^2$, or even $s = \frac{1}{2}at^n$, where n could be any number: the most we could do would be to say that the two series were perfectly (or perhaps imperfectly) correlated. And the differential calculus would be quite inapplicable.

In psychology, therefore, a far more general treatment seems necessary than that proposed by Dr. Taylor; and we suggest that this is provided by the theory of sets and operations on sets. Here it is sufficient to say that this mode of approach leads to a model using linear point sets in a multi-dimensional space of the type known as Hilbertian space. Under certain conditions, which can often be secured if the investigation is appropriately designed, each of the dependent variables can then be expanded as a weighted average of the independent variables. Thus,

$$s_1 = f(x_1, x_2, \ldots),$$
$$s_2 = f(x_1, x_2, \ldots),$$
$$\ldots \quad \ldots \quad \ldots,$$

reduces to

$$s_1 = a_{11}x_1 + a_{12}x_2 + \ldots,$$
$$s_2 = a_{21}x_1 + a_{22}x_2 + \ldots,$$
$$\ldots \quad \ldots \quad \ldots \quad \ldots,$$

and the problems arising can be readily handled by means of matrix algebra or

some equivalent procedure. Under other conditions the dependent variables must be expanded as a power series[1].

The simplest and most familiar instance of this type of analysis involves the determination of the eigen-values (latent roots) and the eigen-vectors (latent vectors) of the matrices that represent the operators; and, in psychological work, is commonly known as factor analysis by weighted summation. This is a branch of multivariate analysis, which in turn is a branch of what is sometimes known as Hilbertian analysis. Thus our reply to Dr. Taylor and to Mr. Willis might be summed up by saying that we believe the most appropriate method of analysis in psychology to be not 'Galileian' or 'Newtonian' analysis but 'Hilbertian'.

Now it is this type of analysis that is chiefly used in quantum theory. The quantities involved are represented, not by space-time functions, but by linear operations in a vector space; and changes in the quantities specifying the state of a given system are represented, not by differential equations, but by changes in the direction of the vector describing the space.

These then are the main resemblances[2] that we had in mind in discussing the analogies between the problems of the psychologist and those of the quantum theorist. But our main reason for insisting on the probabilistic or statistical mode of treatment, which Dr. Taylor and Mr. Willis so severely censure, arises out of recent work on the nature of cortical activity. This has been most clearly expressed and demonstrated by Dr. Sholl: as he puts it, "whether the cortex is studied by the anatomist, the physiologist, or the psychologist, the model employed should be based on the concept of probability, and discussed in a statistical language"[3].

The Meaning of Cause. Mr. Willis's final argument seems to reduce itself to the twofold statement that all scientific laws are in point of fact 'causal laws' and that, if an alleged law is not 'causal', it cannot be genuinely 'scientific'. To

[1] See Burt, *Factors of the Mind*, chaps. III and IV. For the statistical psychologist perhaps the most convenient introduction to the theory of sets is to be found in the opening chapters of Cramér's *Mathematical Methods of Statistics*. Incidentally, since Mr. Willis thinks the reference to Fourier analysis and Fourier integrals irrelevant, it may be noted that they play an important part in the modern theory of the characteristic functions of distributions and the generating functions of probabilities (cf. Cramér, *op. cit.*, pp. 89f.).

What we have said does not necessarily exclude the older models altogether. For example, in research on the effects of conditions which change with change of time (e.g. mental development, learning, fatigue, etc.), and generally in what one of us has termed 'psychodynamics', time appears as a dependent variable with its own definite metric; and it is then still possible to express, at least in an approximate fashion, the 'trajectories' involved by means of differential equations and what Mr. Willis would call 'Galileian analysis'. But even so the method is likely to yield, not the 'universal laws' that Dr. Taylor demands, but only a first crude approximation to the descriptions required; and it is certainly not the typical method of current psychological research.

[2] Most of them have been noted incidentally by quantum physicists themselves, e.g. by D. Bohm in *Quantum Theory*, N. Bohr in *The Atomic Theory and the Description of Nature*, and E. Schrödinger in *What is Life?*; but they seem to have been overlooked by psychologists themselves.

[3] D. A. Sholl, *The Organization of the Cerebral Cortex*, 1956, *ad fin.*

deal with such a statement, however, is by no means easy when we are never told what particular meaning the writer attaches to the word ' causal '[1].

The term ' cause ' may imply one or more of three conditions. As a rule, when scientists themselves speak of a ' causal relation ', all that their arguments really require is (i) the far more general and abstract relation of ' ground and consequent ' (the so-called ' if-then relation '). In everyday parlance, however, when the ordinary person refers to causation, he has in mind something much more concrete than this: nearly always he implies (ii) a relation of temporal sequence; the effect is thought of as coming after its cause, and this means that the dimension of time is regarded as having an irreversible direction. But in addition it often seems also to be generally assumed that, as Mr. Willis puts it, (iii) " the causes must be real causes ": i.e. they must have concrete existence: a cause apparently is thought of as a kind of ' force ', which operates by immediate spatial and temporal contact and in some way compels the effect to occur.

Elsewhere he insists that the cause itself must be ' verifiable '. But no one can verify a cause as such—at least in the material world[2]: for no one has observed, or could observe, the actual process of causation. Thus, even if a cause was in fact a *vera causa*, we could never directly demonstrate it. A cause therefore is always an element in a hypothesis; and, if our analysis of the usual meaning of the word is correct, the hypothesis must embody three distinct assumptions: (i) a relation of ground and consequent; (ii) a relation of immediate succession; and (iii) a unique power or capacity whereby the cause always forces the effect. Now our general contention was that in a pure science the first ingredient alone is essential for whatever inferences or predictions such a science attempts to make, that the second may or may not appear, and that the third has long ago been discarded, as belonging more to the anthropomorphic notions of magic than to science. So far as we can see, Mr. Willis's arguments prove nothing more than the need for the first assumption[3].

To show that ' causal laws ' are indispensable, Mr. Willis appeals to the physical sciences: and he maintains that psychology should model itself upon these sciences. The laws of physics, he tells us, are " expressed by the various equations of dynamics, which state in quantitative form the precise relations

[1] Mr. Willis says " cause means whatever enables us to make a prediction ". But he can scarcely intend this as a complete definition: to do so would be to accept the tacit assumption of the college don, who declared: " It is time we changed the common room barometer; it no longer has any effect on the weather ".

[2] This qualification is added to allow for the fact that some psychologists claim to be aware of a direct causal relation in the case of effective conation. Mr. Willis would obviously not accept such introspective reports at their face value, and in any case he is referring throughout his argument to material processes (where material includes physiological).

[3] Mr. Willis cites Russell as holding that ' causal lines ' are a necessary postulate for scientific methodology. But Russell in the passage cited uses ' causal ' in a rather Pickwickian sense. " A causal law, as I shall use the term ", he writes, " may be defined as a principle in virtue of which, given sufficient data about certain regions in space-time, it is possible to infer something about certain other regions: the inference may be only probable " (*op cit.*, p. 326). He expressly distinguishes his conception from the plain man's notion of a cause, which, as he says, " is primitive and unscientific ".

between cause and effect''. But Kepler's well-known 'laws' of planetary motion make no mention of causes. Nor do most of the 'equations of dynamics'. However, let us look more closely at a specific instance, e.g. what the textbooks sometimes call 'Galileo's law of the pendulum'. This states that, ''if l is the length of a freely oscillating pendulum, then the number of oscillations per second will be $\sqrt{(g/l)} \div 2\pi$''. Now here plainly the 'length' is in no sense a *cause* of the oscillations; nor does the word 'will' refer to a future effect that comes *after* the cause. All that the proposition states is an 'if-then' relation. When, for instance, Euclid declares that ''*if* two angles of a triangle are equal, *then* the sides opposite to them *will* also be equal'', he does not mean that the equality of the sides is a subsequent effect caused by the present equality of the angles: all that he is affirming is a conditional proposition asserting that a logical consequent depends on and follows from a logically antecedent condition.

Mr. Willis, however, may reply that the real cause is the 'force' of gravitation. To that we may answer, first, that in the basic equations of modern dynamics 'force' seldom appears as an explicit term, and when it does it is no more than a verbal synonym for the quantity $F = mg$, where F in turn is merely an algebraic symbol in a mathematical formula. 'Force' in the modern textbook does not mean something that 'forces' the pendulum to oscillate or the moon to revolve in an elliptical orbit. And once again the equations which form the premises of such an argument merely state the 'grounds' for the logical 'consequents'; and the necessity is a logical necessity, not the necessity of physical compulsion.

In any case the modern scientist does not (as Mr. Willis claims) seek to ''depict the world as it really is''. What he describes is not a flesh and blood reality, but merely a skeleton. Physics is palpably based on a drastic set of abstractions. Even Galileo expressly warns his readers that the world with which he is concerned is a world of abstract objects, that are characterized solely by shape and number, and do nothing except exist or move in space and time[1]: to these characteristics the modern physicist would add mass[2]. All 'secondary qualities' he, like Galileo,

[1] *Il Saggiatore*, Quest 48. Because the modern physicist, like Galileo, '' banishes '' secondary qualities, such as colour, sound, and smell, it does not follow, as Mr. Willis apparently supposes, that the psychologist must also exclude them from his field of study. And certainly it is incorrect to suppose that physics banishes all '' qualitative states ''. Physics recognizes many differences in quality, which are in no way psychological. If a cubic inch of ice, and the same amount of water, and again of steam, are placed at the bottom of a beaker, their behaviour will exhibit qualitative differences. Certainly these differences may be interpreted as 'effects' of quantitative differences, viz., of differences in the relative speed of the random movements of molecules. But the fact that the solid block of ice retains its shape, that the liquid conforms to the shape of the vessel, and that the vapour does neither, is itself a qualitative difference of 'state', and cannot be regarded simply as the effect of the way we perceive such things. Moreover, the phrases 'melting point', 'boiling point', etc., remind us that the change of state is relatively abrupt. And, when the physicist speaks of 'states of matter' and 'states of the atom', why should not the psychologist speak of 'states of mind', if he finds the concept helpful? Certainly it is hardly correct to cite physics as a model science which has banished the notion of states.

[2] Salviati indeed asks '' whether there be not in movable objects another intrinsic quality which makes it averse to motion '' (i.e. inertial mass). *Due Massimi Sistemi*, Giorn. II*a*.

would relegate to psychology. An atom has no colour; an electron makes no sound; a neutron has no taste; a proton does not even smell. And he therefore readily admits that what he describes is not the real world, or real entities behind the apparent world, but merely a symbolic and conventional model constructed on a schematic basis, and deliberately simplified so that rigorous argument shall be possible. To object to factor analysis because the factors are 'abstract' is to object to the very technique which renders science possible. Theoretical science is by its own nature abstract. We might just as well criticize a map because it doesn't look like the landscape.

Whenever, in ordinary dynamics, a physicist cites what Mr. Willis would call a 'causal law', and applies it to successive conditions or 'states', all he means is simply this: in discussing a finite closed system, we are to assume that the relation between the two successive states of the system will be (*a*) unique (i.e. one will always entail the other) and (*b*) invariable (i.e. this entailment will always hold good). In such a case both the cause and the consequence must be states of the *whole* system, not the conspicuous partial conditions that the practical man would generally single out as 'causes' and 'effects'. But, though completely closed systems may appear in our theoretical model, they are never found in nature. They have the fictitious character of idealizations. The reference to *successive* states occurs simply because in the differential equations the derivative of time appears as an independent variable. And the fact that the relation specified is an invariable relation (i.e. does not itself change with time) arises simply because time itself does not appear explicitly in the equations. In all such cases the necessity is a formal necessity; and the relation specified is symmetrical. Hence the law enables us not merely to predict a future state but also to postdict a previous state.

When we turn to thermodynamics, we certainly find that the 'second law' now imposes a direction on 'time's arrow'. Here we can no longer make retrospective inferences. Yet even this does not introduce causality as a 'real' or objective concept. The second law is a statistical law. Hence it does not imply exact determination. But in any case our interest in time is practical rather than theoretical; the importance of time relates rather to our aims and desires than to the part it plays in abstract knowledge. Finally, when we get down to microphysical phenomena, the assumption of continuity disappears altogether. Lewin's statement (cited by Mr. Willis) that "in modern physics... classifications have been entirely replaced by continuous gradations" (*loc. cit.*, p. 4) is incorrect. The very term 'quantum' implies that 'action' is discontinuous. Here assuredly it becomes difficult for the most ardent causalist to bring in the notion of causality.

On the other hand, when we have to deal with the everyday world of human affairs as we do in the applied or practical sciences (e.g. clinical, educational, or industrial psychology as distinct from pure or general psychology), the position is different; there we see no reason why the scientific investigator should not, if he wishes, use everyday concepts and postulates, such as causation or even freewill. But this is another story. As Mr. Willis himself has emphasized, the

problems and the methods of the applied sciences often differ widely from those of the theoretical; and it is solely with the latter that he is concerned.

POSTSCRIPT

In his reply (pp. 133–136 below) to the foregoing criticisms Dr. Taylor seems appreciably to have modified his original views. Although he still insists on the ' inconclusive character ' of the ' statistical approach ', he no longer wishes to exclude it altogether. On the other hand, he apparently supposes that we ourselves would exclude every approach that is *not* statistical. Accordingly his main argument is now devoted to demonstrating " the need for non-statistical checks " and the elimination of " logical errors ". With that we heartily agree. However, even if the method of verification is statistical, the laws so verified do not (he says) become statistical laws; and he still maintains that no genuinely ' scientific law ' can be statistical. Such a law is, or ought to be, exact, universal, and rigidly deterministic; and once again he cites Newton's law of gravitation as an ideal instance.

Now, we should be quite prepared, for the purpose of hypothetical deductions, to include in the *hypothetical model* a ' law ' which, like that of Newton, is stated in the form of an ' exact and universal ' proposition. But we should hold that the *applicability* of such a model to the concrete phenomena to be explained is likely to be approximate rather than exact, and probable rather than universal or unconditional. At one stage of his argument Dr. Taylor himself seems almost inclined to accept this point of view, since he remarks that " the *probability* [of such a law] is increased with each deduction that is confirmed "—which was virtually our point in emphasizing the " consilience of inductions " (p. 116).

But, in holding up, as a model for the psychologist to imitate, Newton's discovery of the ' universal law ' that governs the movement of the heavenly bodies, Dr. Taylor, as it seems to us, is overlooking the marked difference in complexity between astronomical phenomena and psychological. The earth, the moon, and the sun happen to be bodies far removed from other causal influences. Hence Newton had comparatively little difficulty in reaching a highly exact estimate for G by considering their spatial relations and velocities; and it is for precisely the same reason that astronomy—at least the astronomy of Newton's day—seems at first sight to be a highly exact science. Chemical phenomena are more complex; and biological phenomena still more so. Hence in those sciences the conclusions are of necessity far from exact.

However, the chief difficulty we find in answering Dr. Taylor's criticisms arises out of his own logical procedure, or rather (we are tempted to say) his lack of it. Like Mr. Willis he follows Lewin in preferring the ' Galileian way of thinking ' to the Aristotelian. And certainly his own method of exposition has something of the literary and rhetorical style of Galileo as contrasted with the dry and formal logic of the typical Aristotelian. That certainly makes his arguments far more readable; but it leaves them far less cogent. This preference for a free and easy style we believe to be prevailing weakness of contemporary psychology as a branch of natural science. And it was largely for this reason that (taking Dr. Taylor's own example—the study of astronomical phenomena) we ventured to draw attention to Scheiner's method of exposition in his disputes with Galileo over the astronomical problems that Dr. Taylor discusses: with such a method the reader knows at once what precisely is the question under discussion, which are the postulates and which the factual premisses, and what is the nature of the reasoning adopted in proving the final conclusions.

Now in Dr. Taylor's contribution we cannot help suspecting that he is primarily concerned with questions that are quite distinct from those which the other contributors are discussing. Ostensibly he seems to be attacking the conception of scientific and statistical method put forward in the previous issue of the journal; but actually all he seems to

do is to state afresh and illustrate anew the hypothetico-deductive procedure regardless of the previous criticisms or of the alternatives suggested. Thus he begins by asserting that the verification of a scientific theory is based on the scheme " If *A*, then *B* ", where " *A* is verified if *B* is found to be true ". Let us translate this argument in the form of a concrete example: " If Aristotle's law of gravity is correct (*A*), then a bullet will drop to the ground more swiftly than a feather (*B*); but the bullet is found to drop more quickly; therefore Aristotle's law is verified ". Now Dr. Taylor, we imagine, would be the first to reject such an inference. Here there is surely a " logical error " which he himself has failed to eliminate.

When the " logical errors " have been removed, we are told, " the first deductions are likely to be of an all-or-none character ". But that seems to imply a return to the Aristotelian (or pseudo-Aristotelian) principle of ' two-valued logic ', namely, " *x* must be either *a* or not-*a*—there must be no *tertium quid* ", in the teeth of all Lewin's objections to such dichotomies. In his previous paper he gives an instance. There he quotes the current controversies about nature versus nurture. His contention is (if we follow him aright) that any given psychological characteristic must be *either* wholly innate *or* wholly the result of experience: for, unless the two are mutually exclusive, no deduction is possible. The statistical psychologist of course maintains that every observable characteristic is the compound resultant of *both* factors. But that, says Dr. Taylor, is tantamount to " the dethronement of logic ".

In the foregoing pages we have endeavoured to prove that Dr. Taylor's nineteenth century picture of the material universe is no longer tenable; and this denial of universal determinism was reached long before the modern quantum theory was developed or accepted: probably the most celebrated and convincing criticisms of the traditional determinism were those of Franz Exner after the commencement of the present century.

However, perhaps the simplest and most accessible discussion of Dr. Taylor's problem is that contained in Professor Schrodinger's recent book on *Science, Theory, and Man*, which appeared after our own paper was written. If Dr. Taylor will refer to the chapter entitled ' What is a Law of Nature ? ' he will find an answer to his arguments from the pen of one of the leading physicists of the present day. Schrodinger expressly discusses the older view, which Dr. Taylor implicitly accepts, and which held that " the law of gravitation is of the absolute type and is perhaps at the basis of everything as *the* primary and fundamental law ". He shows that a survey of the so-called " laws of nature " indicates that any such law " loses its exact validity in proportion as the number of cooperating individual processes diminishes "; and points out that " in no case is the number of single atoms that must cooperate so vast as in the case of gravitational phenomena—a convincing proof of the essentially statistical character of such laws ".

We do not question the value or validity of the interesting research on " the behavioural basis of perception " which Dr. Taylor has carried out and described in some detail in the second section of his paper (pp. 134f.). But we can see little in it that is incompatible with the view of scientific and statistical method developed by Burt in the previous issue of this journal, and nothing to rebut the criticisms there raised. Nor is any further defence offered for Taylor's original contention that in psychology no " genuine scientific theories are likely to be established by statistical investigations ", or for his former attack on " the use of all the apparatus for testing significance, analysing variance, and the like " as a mere " cloak for intellectual sterility ". We therefore venture to think that Dr. Taylor's reply would be far more convincing were he to state quite explicitly which of the criticisms he now accepts, which of them he repudiates and why, and which (if any) are based on a misunderstanding of his real position,

W. L. Gregory.

Vol. XI Part II The British Journal of Statistical Psychology November 1958

SCIENTIFIC METHOD IN PSYCHOLOGY: III

By G. W. Pilkington
University of Sheffield

I. *The Specificity of Scientific Explanations.* II. *The Demand for Quantification.* III. *The Objections to Dogmatic Methodism.*

I. The Specificity of Scientific Explanations

Mr. Willis and I probably agree more than he thinks; but to convince him of this I shall have to expand my argument from the brief and abstract form that appeared in the *Bulletin*. In suggesting that there is " no *one* eternally correct method of explanation for psychology," I did not for a moment wish to " fly in the face of the views expressed by nearly every writer on scientific method ". For many purposes, it is indeed both convenient and important to talk of generally accepted standards of what constitutes scientific proof and explanation[1]: in this sense there may be a " general logical pattern " common to all the sciences. And, if psychology is to be called a science, it should conform to these logical standards.

But I would argue that what can sensibly be said about scientific explanation *in general* (i.e. so that it applies to *all* those spheres of enquiry we commonly call scientific) is much more limited than is commonly supposed. In this broader sense, scientific explanation consists of statements (*a*) which can be confirmed intersubjectively; (*b*) which specify, or enable us to specify by some conceptual device, relations of either a probabilistic or a deterministic kind between observable phenomena; and (*c*) which are the result of systematic, and not casual, observation—including experimentation if possible, though not of necessity. Requirement (*a*) rules out metaphysical, theological, moral, aesthetic, and other non-empirical statements[2]; requirement (*b*) shows that " the fundamental concept for science is thus that of scientific law, and the fundamental aim of a science the establishment of such laws "[3]; and requirement (*c*) shows that scientific statements, if true, have undergone a social testing procedure, both lengthy and rigorous.

[1] For especially good examples in addition to those cited by Mr Willis, see Hospers, J., ' On explanation ', *J. Philos.*, **43** 1946, pp. 337–356; Braithwaite, R. B., *Scientific Explanation*, 1953; and Popper, K., *The Open Society and Its Enemies*, 1945, Chap. XXV., especially pp. 261f.

[2] Cf. Braithwaite, *op. cit.*, p. 2: " A scientific hypothesis . . . is an empirical proposition in the sense that it is testable by experience; experience is relevant to the question whether or not the hypothesis is true."

[3] Braithwaite, *ibid.* Later (p. 9) he continues: " What is a scientific law? The one thing upon which everyone agrees is that it always includes a generalization, i.e. a proposition asserting a universal connection between properties: it includes a proposition stating that *every* event or thing of a certain sort either has a certain property, or stands in certain relations to other events or things having certain properties,"

A stricter definition would immediately exclude from the domain of science spheres of enquiry that nevertheless form perfectly legitimate and respected parts of Faculties of Science in the Universities. If, for example, we insist that ordinary empirical generalizations or tendency statements (cf. (*b*) above) do not constitute genuine scientific explanations, or that they do so only when they are conceptually unified so as to be deducible from some higher level set of theoretical statements[1], then clearly much of what ordinarily passes for science would no longer be classed as science. Obvious examples include the biological and medical sciences, economics, geography, geology and meteorology, none of which, to my knowledge at least, exhibits a definite hypothetico-deductive system.

II. The Demand for Quantification

The same applies to the requirement of quantification on which Mr. Willis insists so strongly. Peters[2], exploring the influence on psychology of Kant's view that an empirical enquiry is scientific only so far as it is mathematical, has pointed to some of the more unfortunate consequences of a too stringent adherence to this dictum. It may, of course, be argued that a science is a *better* science, or a more advanced science, if it provides a conceptual device for deducing fully quantified, empirical generalizations; and I myself should agree. But to *define* science in such a strict way is to commit the mistake of applying to young or immature sciences standards appropriate only to the advanced or well developed sciences[3]. More specifically, it is to treat psychology as if it were physics, a mistake to which many philosophical commentators have drawn attention[4]. Yet all of us are aware that psychology is concerned with extremely complex phenomena—the product of many heterogeneous variables, which are difficult to manipulate experimentally; whereas physics, at least classical physics, deals with simpler phenomena[5] which can be more easily manipulated experimentally, and formulated mathematically in ways more obvious than the phenomena of human experience and behaviour.

III. The Objections to Dogmatic Methodism

' Dogmatic methodism ' (this useful phrase originated with Peters[6], not with me) may take many forms. In my paper the varieties to which I chiefly

[1] This view has been put to psychologists with greatest force by Hull: cf. esp. Hull, C.L., *Principles of Behaviour*, New York: Appleton-Century-Crofts, 1943, pp. 1–31, and Hull, C.L. *et al. Methematico-deductive Theory of Rote Learning*, 1940, pp. 1–13.

[2] Peters, R. S. (Ed.), *Brett's History of Psychology*, 1953, pp. 507–509.

[3] Braithwaite, *op. cit.*, p. 1. " If the science is in a highly developed stage, as in physics, many special laws will appear as logical consequences of a small number of highly general laws expressed in a very sophisticated manner: if the science is in an early stage—what is sometimes called its ' natural history ' stage—they may be merely generalizations involved in classifying things into various classes ".

[4] E.g. Peters, *op. cit.*, pp. 719f.; and especially Farrell, B.A., Braithwaite, M. M., and Mace, C. A., Symposium on causal laws in psychology, *Proc. Arist. Soc. Supp.*, XXIII, 1949, pp. 31–68.

[5] As Gregory has also pointed out, *Brit. J. Statist. Psychol.*, VI, 1953, p. 108.

[6] *Op. cit.*, p. 716. The whole of Peters' last chapter on ' Twentieth-Century Trends ' (pp. (658–725) provides instructive reading on this question.

objected were, so far as psychology is concerned, the following: (*a*) The selection of some particular class of empirical variable as primary causes, e.g. physiological and neurophysiological variables, as suggested by those who believe that all explanation in psychology must ultimately be physiological[1]; (*b*) the elevation to explanatory primacy of some particular type of generalization, e.g. one couched in teleological terms, as advocated by McDougall; (*c*) the rejection, or at least the apparent rejection, of all the ordinarily accepted canons of scientific explanation, and the substitution of some special and unusual type of explanation alleged to be peculiar to psychology, as urged by the proponents of ' understanding ' psychology; and (*d*) the contention that only some particular kind of formal structure of propositions, from which lower-level empirical generalizations are deducible, constitutes a proper scientific explanation, e.g. the ' hypothetico-deductive ' system eulogized by Hull and Eysenck.

I have already tried to indicate some of the defects in the position held by these last two writers. But my strictures were reserved for hypothetico-deductive proofs as expounded in their somewhat narrow and dogmatic sense. There is a wider and far less tendentious sense in which the phrase is often and perhaps not quite accurately used; and to this I should make no objection[2]. Thus I should certainly agree that it is " almost a platitude to say that every science proceeds, more or less explicitly, by thinking of hypotheses, of greater or lesser generality, from which particular consequences are deduced which can be tested by observation and experiment "[3]. In this sense the phrase ' hypothetico-deductive ' seems simply to cover scientific explanation in general, with attention specially drawn to the intimate connection between deduction and the observations required for testing hypotheses. Nevertheless, I would emphasize that quite isolated scientific hypotheses, not conceptually embedded in any hypothetico-deductive system in the narrower sense, and of a relatively low level of generality, may yield observationally testable consequences; indeed, this *must* be so in the case of any explanation in which recourse is had to a general principle other than a mere generalization of the explicandum. Of course, a typical hypothetico-deductive system is almost always very fertile; but a simple and isolated scientific hypothesis may also bear deductive progeny, though inevitably its offspring will be fewer. And I see no reason why the title of ' scientific explanation ' should not be accorded to both, even though the hypothetico-deductive system is awarded the methodological laurels.

Mr. Willis does not sufficiently distinguish between the wider and narrower senses of the term ' hypothetico-deductive method '. In a very general sense almost all our problem-solving involves setting up hypotheses and making

[1] E.g. Pratt, C. C., *The Logic of Modern Psychology*, 1939; Hebb, D. O., *The Organization of Behaviour*, 1949; Klein, G. S., and Krech, D., ' The problem of personality and its theory ', and Hebb, D. O., ' The role of neurological ideas in psychology ', both republished in Krech, D. and Klein, G. S. (Eds.), *Theoretical Models and Personality Theory*, 1952.

[2] Mr. Willis mentions Boring's contribution to *Foundations of Psychology*. But I think examination shows that Boring is using the phrase in the wider and looser sense.

[3] Braithwaite, *op. cit.*, p. ix.

deductions from them. But in modern psychology, the phrase usually refers to the abstract and highly formalized procedure advocated by such methodologists as Hull and his disciples. Many would certainly question whether this particular form of it is *the* method for psychology. However, Mr. Willis is not apparently urging us to adopt the paralysing procedure preached and practised by Hull: rather he uses this fashionable phrase as a stick to castigate his real enemies—statistical or probabilistic laws. Statistical laws, we learn, can never be 'genuine scientific laws'; and therefore the psychologist's task is to formulate fully deterministic, or 'causal', laws. But this *is* dogmatic methodism, naked and unashamed. In the biological and social sciences as well as in educational and social psychology, statistical laws occupy a central place; and it is surely beside the point to argue that they can never be 'genuine' or 'proper' laws, just because they are not typical of classical physics. At the same time, I do not wish to imply that in psychology *all* laws should of necessity be statistical: that would be dogmatic methodism of an opposite kind. In either case the attempt to lay down *a priori* the form that the relations between phenomena *must* take is equally futile: our job is to find out precisely what the relations are.

I would agree, then, with Mr. Willis that there is a logic of explaining behaviour and experience; but I hold that it is dangerous to insist on anything more specific. Certainly a methodological analysis of psychology is not to be written off as worthless: on the contrary, it should be done better, that is, with a fuller regard to the explanations actually used in psychology. Otherwise, we may be led to condemn in advance the efforts of colleagues who are working with ideals of explanation different from our own in spite of the fact that in the end their results may prove to be just as valuable.

Vol. XI Part II | The British Journal of Statistical Psychology | November 1958

SCIENTIFIC METHOD IN PSYCHOLOGY: IV

By J. G. Taylor
University of Cape Town

I. *Statistical and Non-Statistical Verification.* II. *An Illustrative Psychological Research.* III. *The Inconclusive Character of the Statistical Approach.*

I. Statistical and Non-Statistical Verification

The verification of a scientific theory starts from a statement in the form: ' if A, then B ', where A is the hypothesis to be tested and B is a deduction from A. It is essential that A be stated with such precision that ambiguity is completely excluded and its implications deduced by strict logical or mathematical process. Assuming this twofold condition to be satisfied, A is verified if B is found to be a true description of the relevant natural phenomena.

If B is a quantitative statement, verification will as a rule require statistical checks, since the measurement of both independent and dependent variables is fallible. The theory so verified does not, however, thereby become a statistical law. Thus Newton's law that the force attracting two bodies of mass M_1 and M_2 at a distance d is equal to GM_1M_2/d^2 does not become a statistical law merely because the value of G is stated as an average plus or minus a probable error. To call it a statistical law would seem to imply that the gravitational constant is in fact a variable quantity. This would be true only if the measurements of M_1, M_2, and d were infallible. But in fact the masses M_1 and M_2 used in experiments designed to give an estimate of G are at the same time exposed to the earth's gravitational pull; and measurements with the Eötvös balance show that this is subject to local variation, which has been attributed to variation in the density of the earth's crust.

Any good theory mediates a large number of deductions, and the probability that it is a true explanation of the natural phenomena to which it refers is increased with each deduction that is confirmed. Thus confidence in Newton's law was increased when Leverrier deduced from it and from the perturbations in the orbit of Uranus that there must be a still more distant planet, Neptune, and Neptune was in fact discovered. As Burt has pointed out (see p. 111), Leverrier also deduced from the advance in the perihelion of Mercury that there must be another planet, or group of planetoids, Vulcan, revolving within the orbit of Mercury. There is no such planet. The logic of the argument is that if A, Newton's law, and B, Vulcan, then C, the advancing perihelion. Then not-B implies not-A. It is therefore argued that Newton's law, having survived all the tests to which it was subjected over more than 200 years, was finally disproved by a single fact—the non-existence of Vulcan. But even so the crucial test would be non-statistical.

II. An Illustrative Psychological Research

To illustrate the distinction between statistical and non-statistical verification I propose to refer to some deductions from my own theory of the behavioural basis of perception. A brief summary of this theory has been published[1]; but the full theory, from which the illustrations are taken, is not yet ready for publication.

The first deduction is that if a person, sitting at a table on which there are a number of objects, tilts his head from side to side while keeping his body erect, and observes the displacements of objects relative to one another, without fixating any object but allowing the eyes to wander freely over the table, he will find that there is a relatively fixed centre of rotation about which all the other objects appear to move. The position of this centre of rotation is such that it can be reached by stretching out the hand at arm's length. If the subject, still sitting at the table, sways his body from side to side there is a new centre of rotation, which can be reached by the outstretched hand with the aid of an extensive forward inclination of the body. If he now stands and sways from foot to foot the centre of rotation has receded still farther, but the theory does not permit of any deduction regarding the distance of the new centre.

Now the chain of reasoning that led to this deduction includes the assumption that the subject, at an early stage in his life, could move his head and arms but could not sit, and at a later stage could sit but was incapable of locomotion. Since this is true of the overwhelming majority of people the deduction can be generalized to the population. Individual differences in health, maternal care, and other factors that might affect the infant's adaptation to his environment are not specifically considered in the theory in its present form; and the distances deduced for the first two centres of rotation are therefore stated as averages. It would seem that in this case verification requires both non-statistical and statistical checks. In so far as the deduction specifies exactly two relatively fixed centres of rotation and a third whose distance is indeterminate, the test is non-statistical. The theory does not permit any deduction regarding the proportion of subjects who will be unable to follow the instruction to allow the eyes to wander freely while observing the relative displacements of objects; consequently there is no statistical check of the proportion of subjects who report three centres of rotation, increasing in distance according as the head, the trunk or the whole body is in motion. On the other hand statistical checks constitute the only possible means of verifying the deduced distances of the first two centres of rotation. This is due not merely to the fallibility of measurements but to the essential character of the deduction itself. It is drawn from a theory that takes no account of possible sources of perturbation in the expected results.

There is, however, nothing to prevent specific sources of perturbation being taken into account, and this, as the next illustration shows, permits a more searching test of the theory. Let a specific subject, viz. J. G. Taylor, suffering

[1] Taylor, J. G. (1957). Towards a science of mind. *Mind*, LXVI, *N.S.*, 434–452.

from the handicap of a club foot, wear spectacles in the form of 20° prisms, with bases to the left, for a continuous period of 5½ hours each day for a fortnight, and let him, during each of those periods, walk for approximately 15 minutes and spend the rest of the time sitting at a table, having meals, and working. The theory leads to quite a number of deductions; but I confine myself to two, concerning the perception of horizontal surfaces below the level of the eyes, specifically the ground on which *S* walks and any table at which he sits. At the beginning of the experiment the effect of the spectacles is that all such surfaces are perceived as sloping down from right to left with an increasing gradient.

Deduction 1 is that at the end of the experiment, when *S* walks on level ground while wearing the spectacles he perceives it as having a level path, about a yard wide, stretching in front of him in the direction in which he is walking. The ground slopes up gently to the right of this path, and down more steeply on the left. The perceived direction of the path is unaffected by lateral movements of the eyes. From the intermittent wearing of the spectacles it can be deduced that the after-effects commonly resulting from continuous wearing will be absent. Consequently if the spectacles are turned round so that the bases of the prisms are to the right, the perceptual distortion will be identical with that experienced at the beginning of the experiment, except that its direction is reversed. This permits direct comparison of the terminal with the initial perceptual experiences.

Deduction 2 is that when *S* sits at a table such that every part of it can be reached from the sitting position, it will be perceived as an oblique plane, sloping down to the left, but with no trace of curvature. In this situation adaptation to gravity presents no problem as it does in walking, but the initially perceived curvature of the table results in behavioural errors that have to be corrected. Thus visual stimulation by a horizontal surface produces, in one and the same subject, two markedly different perceptual structures.

I performed this experiment in Innsbruck with the collaboration of Dr. Ivo Kohler. The results indicated were obtained before the deductions were made (the theory not then having reached its present stage of elaboration), but that does not affect the issue. The point I wish to make now is that I could not generalize these deductions to the population, since the specific values substituted for the general terms of the theory include my own instability on my feet and my own everyday habits of life. In this case a statistical test is ruled out by the specificity of the substituted values, and verification requires that the deduced results be confirmed not within certain confidence limits but exactly. The exact width of the path in deduction 1 is not critical but its slope is, and so is its direction relative to the median plane of the body. If it had not been level, or if it had moved with the eyes, the theory would have been proved false. Similarly, the slope of the table was not critical, but if it had not been flat, again the theory would have collapsed.

To sum up, I might adapt the language of the grave-digger in *Hamlet*, and say that verification has three branches—to check, to confirm, and to verify. When the logical structure of a theory is in question, checking will in general be

non-statistical, since, if the logic is at fault, the theory will sooner or later generate deductions that are manifestly absurd. When the logical errors have been eliminated, the first deductions are likely to be of an all-or-none character. They will take the form that the system possesses such-and-such a property : my own theory, for example, mediates the deduction that in a normal subject a vertical line is perceived as vertical despite variation in the angle at which its image falls on the retina. In such a case confirmation is non-statistical. Subsequent deductions are more likely to be expressed in a quantitative form. Verification must then of necessity be by statistical tests.

III. The Inconclusive Character of the Statistical Approach

This appears to have been the order of development in physics; but in psychology statistical tests are regularly used for the ' verification' of hypothesis that have apparently been subjected to nothing but a most cursory test of logical consistency. To take but one example, it is commonly believed that highly valued objects are perceived as being larger than objects of identical size that are despised. Typical of the experiments on which this hypothesis is based is one by Beams[1], who asked his subjects to adjust the size of a projected colour photograph of an object seen a few seconds earlier to the remembered size of the object. Most subjects made the picture too large if it represented a valued object and too small if the object was disliked. Statistical tests show that there is a highly significant difference between the sizes of the photographs that are matched with valued and with despised objects. But this does not verify the hypothesis that the *perceived size* of an object is an increasing function of the value attached to it, since that hypothesis is in fact logically nonsensical. For if an object is perceived as larger than life size, this will also apply to a realistic photograph of it, and the matched size should therefore agree with the true size. The hypothesis thus generates a deduction that denies the phenomena it purports to explain. In the light of this argument Botha[2] showed in quick succession photographs, identical in size, of liked and disliked foods, and asked her subjects to indicate which, if any, appeared to be larger. The answers showed that when the subjects not merely declared that they disliked a given food but habitually endeavoured to avoid having to eat it, it was perceived as being distinctly *larger* than a preferred food.

If statistical methods can ' verify ' absurd hypotheses, there is surely a good case for my plea that the logic of verification should be considered in its entirety, and that the need for non-statistical checks should be universally recognized.

[1] Beams, H. L. (1954). Affectivity as a factor in the apparent size of pictured food objects. *J. Exp. Psychol.*, XLVII, 197–200.

[2] Botha, E. (1956). A study of the effect of preference on the perception of size. *J. Social Res.*, VII, 49–57.

Vol. XI
Part II

The British Journal of Statistical Psychology

November
1958

STUDIES IN THE DEVELOPMENT OF REASONING IN SCHOOL CHILDREN

I. General Methods and Results

By Dorothy Wheeler
Psychological Department, University College, London

Summary. Earlier researches produced a number of conflicting hypotheses as to the nature and development of the reasoning processes; but of late the whole problem has been singularly neglected. The following investigations attempt to decide between the rival theories by the combined use of experimental, statistical, and introspective procedures. The main conclusions are: (i) that the solution of problems by reasoning involves, in addition to the general factor of intelligence, a number of specific factors (e.g. factors for the apprehension of relations, for the combination of relations, and for both analytic and intuitive procedures); (ii) that, contrary to the view put forward by Hall and Piaget, most of the elementary schemata necessary for valid reasoning are already within the capacity of the average child of seven, so that subsequent development consists chiefly in a more experienced use of these capacities and in an increasing complexity of the problems which the child is able to solve.

I. Problem

Need for Renewed Investigations. It is strange that, at a time when there has been a rapidly increasing need to select and educate those who are specially suited for training in scientific work, the psychological study of the processes underlying scientific reasoning has been so widely neglected. For the past thirty years the interest of psychologists has been focused chiefly on the study of the emotional and social factors in mental development ; and earlier investigations on cognitive processes have commonly been dismissed as over-intellectualistic. During the same period, as numerous publications both technical and popular testify, logicians and methodologists have been busy introducing revolutionary changes in the formal analysis of scientific reasoning : so that the conclusions of previous psychologists, who based their views, their tests, and their experiments mainly on the traditional logic, are now largely misleading if not wholly out of date[1].

[1] There is no need to discuss the views of earlier writers in detail. Bain's textbook of psychology (*The Senses and the Intellect*) devotes considerable space to reasoning; but his account contains little more than what he says about induction and deduction in his textbook of logic; he criticizes in some detail the theories of older writers—Reid who had postulated a ' faculty ' of reasoning, Hamilton who postulated a ' faculty of relations ', and Bailey who classified abilities into a hierarchical scheme containing four ' genera ', the highest of which was Reasoning with subordinate ' species ' included in the ' genus '; and he contends that reasoning is essentially a form of association—association by similarity and what he calls ' constructive association '. Sully's textbook (*The Human Mind*), the standard work until it was replaced by Stout, devotes 30 pages to ' Reasoning ' on somewhat similar lines. In contrast to all this, the textbooks of today often omit it altogether: in Boring, Langfeld, and Weld's *Foundations of Psychology*, there is no reference to the process at all, and the word finds no place in the index.

Previous Work. In this country the earliest experimental studies of the subject appear to have been those undertaken by Burt and his various research students, largely with a view to solving problems of educational method. At Oxford and Liverpool, in cooperation with J. C. Flugel, R. C. Moore, and J. L. Corkhill [4, and refs.], several factorial researches were carried out between 1906 and 1913. These laid the foundation for his well-known scale of reasoning tests [5]. At Cambridge a year or two later he took up the problem afresh in collaboration with Dr. Susan Isaacs ; and, owing to the influence of Russell who was then lecturer at Trinity, began to substitute the ' modern ' logic of Cambridge for the traditional logic of Oxford as a basis for their experiments. The collaboration was resumed when Dr. Isaacs joined the staff of the London Day Training College (now the Institute of Education) and later of University College. The outcome was a fresh series of researches, first on tests of reasoning as a method of testing intelligence [6, 22] and secondly on the factors involved [8, 9, 10, 12, 13, 15].

Meanwhile, Spearman had developed an alternative factorial interpretation (*Abilities of Man*, especially pp. 172f.), based on a vast number of investigations on school children, which led him to reject the views put forward by Burt. " Reasoning, he says, depends *solely* on G; *no* other factors are involved."

In America, Thurstone's conclusions differed both from those of Burt and those of Spearman. In his studies of ' primary mental abilities ' he isolated a specific ' factor for reasoning ', and subsequently distinguished two such factors, for inductive and deductive reasoning respectively, but found no factor for general intelligence [18, 19]. More recently still, J. P. Guilford and his fellow workers have published a series of systematically planned investigations revealing (amongst others) a factor which he calls ' general reasoning '. Earlier results suggested three possible hypotheses as to its nature, viz. (*a*) ' defining problems ', (*b*) ' handling complicated procedures ', (*c*) ' trial and error manipulation ' ; and, using five ' reference factors ', he eventually concluded that " the first hypothesis—' defining problems '—seems to fit the factor identified as ' general reasoning ' better than the other two " ([31] ; see also the tabular schemes at the end of his contribution to the Paris Colloquium on factor analysis [33, pp. 71–2]).

To the British investigator the defect of these American analyses lies in their failure to recognize the influence of ' general intelligence '. However, Holzinger and his colleagues on re-analysing Thurstone's data reached exactly the opposite result ; and in their own researches found no clear evidence for group factors for reasoning : instead " these disappeared into the general factor ". Their final conclusion is that the general factor of intelligence is itself " likely to consist primarily of reasoning ability " [20, 21]—a hypothesis which Wolfle later endorsed in his general survey [23, p. 33].

Theoretical and Practical Problems. The earlier investigations of Burt and his co-workers had a twofold aim, " first to ascertain the mode in which the

reasoning capacities of children develop from year to year, and secondly, by means of tests for these capacities, to estimate the general level of intelligence among older and brighter pupils " [5]. The tests used consisted partly of graded verbal tests and partly of more concrete tests of a performance type ; and, according to the practice of those days, the quantitative measurements were supplemented by qualitative records of the various procedures, comments, and introspective observations of the children, in order to gain light on the nature of the factors involved. In later researches it was the second of these problems—the practical value of reasoning tests for measuring intelligence—which engrossed the attention of most British investigators, and the theoretical problems were almost wholly neglected[1]. Nevertheless, as in every branch of science, a prior study of the underlying theoretical issues is almost essential before efficient applications can be secured. It is impossible to know what we are really measuring when we insert problems in reasoning into our booklets of intelligence tests, unless we have first discovered whether or not there is a specific factor (or factors) for reasoning over and above that for general intelligence, and, if so, at what ages they first emerge or mature. These are the two broad problems with which my own investigations have been chiefly concerned. But, to justify the special aspects and the special methods on which I myself have concentrated, it will be desirable first of all to review quite briefly the main results of previous workers, since these will indicate more clearly the chief alternative hypotheses between which a decision has to be sought.

Conclusions of Previous Workers. According to the theory put forward by Burt, supported both by objective and by introspective evidence, " reasoning appears to be a form of mental association in which the nature of the associations is itself explicitly apprehended : . . . . in short, reasoning consists essentially in the apprehension and application of what are technically termed relations ", that is (if we take as typical its fully developed form) " in the correct understanding of propositional functions and the correct employment of propositional operations ". The special test devised to demonstrate and illustrate this twofold theory was one that he called the ' Analogies Test ' ; a correct solution of such a problem required " (i) the perception, implicit or explicit, of a specific relation and (ii) the reconstruction of an analogous relation

[1] Abroad, however, the position was reversed. Professor Piaget, for example, relates that towards the end of 1919, when he went to Paris to study psychology at the laboratory founded by Binet and Simon, Dr. Simon suggested he should " standardize Burt's reasoning tests on Parisian children ". But he himself could work up no enthusiasm for studying their diagnostic value by just " comparing the number of successes and failures ", and thought it " more interesting to find the *causes* for the failures, and so analyse the verbal reasoning of normal children "—a task that he took up " without Dr. Simon being quite aware of what I was up to ". During subsequent years, Piaget and his fellow workers have been among those who have carried out the most sustained series of researches on the thought processes, rational and non-rational, of children from the earliest ages onwards (cf. J. Piaget, *The Psychology of Intelligence*, 1950, and refs., and *A History of Psychology in Autobiography*, IV, p. 244).

by ' relative suggestion ' " : or, to express the inference in symbolic form, " given $f(a, b) \equiv f(x, y)$, then $b = F(a)$, and therefore $y = F(x)$ " [5, p. 101][1].

The theory was later adopted by Spearman and incorporated in his doctrine of the noegenetic principles of cognition, where the two processes are re-named the ' eduction of relations ' and the ' eduction of correlates '. Spearman, however, regards them as essential characteristics of cognitive processes generally, and declines to recognize any kind of special reasoning ability.

The simplest tests (Opposites, Similarities, Part-and-Whole, Cause-and-Effect, etc.) involve the perception of single relations only ; the scheme of the Analogies test involves two relations ; ' Syllogisms ' three. And Burt's scales of reasoning tests are constructed in this way by taking all the basic forms of inference, and, with the aid of a symbolic notation, building them up into systems of increasing complexity and variety, and finally embodying each in some simple concrete example, both verbal and non-verbal [5, 6, 22].

The conclusions which he and his fellow workers drew from the results obtained are in sharp conflict with the older views. Thus, in regard to the nature of the reasoning process, they contend that reasoning depends, not on a single factor or faculty, such as ' intelligence ' or a ' primary ability for reasoning ', but upon a number of factors both general and specific [28]. And as regards the development of reasoning they vigorously deny that reasoning is an ability which matures late in the child's development : on the contrary, they argue, " all the elementary mechanisms essential to logical reasoning are present in the average child before he leaves the infants' department, i.e., at least by the mental age of seven " ; and they equally deny any qualitative changes in the development of reasoning ability, such as those alleged by many continental psychologists, notably Piaget[2].

It will be seen that on all the issues thus raised there are sharply opposed hypotheses between which no final decision seems as yet to have been reached. It was the purpose of the experiments to be described in this and

[1] In modern logic a ' property ' is defined as a ' one-place situational function ' and a ' relation ' as a ' two-or-more-place situational function ' and each is expressed by the corresponding linguistic function (cf. [25], pp. 82f.). The view that we can directly perceive relations (as well as mere sensory contents) was taken over from Spencer and James. The principle of ' proportionate ', ' relational ', or ' relative ' suggestion (nowadays known as ' eduction of correlates ') was due to Thomas Brown and developed more particularly by Stout, as part of his theory of ' noetic synthesis ' (*Mind*, XVI, pp. 50f.; *Analytic Psychology*, II, pp. 78f.; cf. also [22] and refs.). Spearman originally held that the fundamental process in ' general intelligence ' was ' sensory discrimination ' and denied the perception of relations; his acceptance of ' noetic principles ' first appears in *The Nature of Intelligence and the Principles of Cognition* (1923, ch. v and vii; see also *The Abilities of Man*, 1927, pp. 165f., 201).

The Analogies test has proved one of the most popular: it appears in Whipple's *Manual* (2nd ed.), in Terman and Merrill's revision of the Binet scale (who describe it as " one of the most satisfactory from every standpoint "), and in most of the group tests used at the present day for the 11+ examination: in all these cases it is regarded as a test of general intelligence rather than of reasoning.

[2] The most succinct summary of the foregoing conclusions is to be found in the joint appendix by Burt and Isaacs in [16] and in Burt's appendix to [11, pp. 254 to 268]: see also S. Isaacs, *Intellectual Growth in Young Children*, 1930, especially pp. 78f.

subsequent papers to confirm or to confute the various alternative theories. In this introductory paper I shall content myself mainly with outlining the experimental methods employed and with briefly summarizing the main conclusions so far reached[1].

II. Methods

Tests. As the investigators quoted continually point out, reasoning is too often treated by psychologists as if it were limited chiefly to deductive processes (to the exclusion of inductive) and conceived and carried out in verbal terms. For purposes of testing, verbal tests of a deductive type have the advantage of being short, speedy, easy to mark, and easy to apply to large numbers. But, for the systematic study of the actual development of the child, it is far more instructive to work with problems of a more practical type, requiring inductive inferences rather than deductive. Such problems have necessarily to be presented to the children individually, particularly if their procedures are to be observed and their introspective explanations obtained. The method is slow and cumbersome and makes any rigorous statistical check more difficult.

Group Tests. In order to secure some general notion of the special factors likely to be involved in processes of reasoning as carried out by children, we began by making a factorial study of results obtained from group tests of the usual type, in verbal and non-verbal form. In selecting the various tests we based our tentative hypothesis mainly on the investigations cited above but partly on more recent researches carried out at London and elsewhere [14, 26, 27]. Our aim was to ascertain how far the various factors discussed by the different writers could be verified or disproved. The problems chosen comprised most of the commoner types of reasoning test (described more fully in [5, 6, and 22], as well as in the unpublished theses just referred to). Supplementary tests for general intelligence, verbal ability, 'fluency', and span of attention were also included. The general nature of each test is sufficiently indicated by its designation in the table (p. 144). On factorizing the tables separately, indications of minor differences between the boys and the girls were discernible, but they were too small to be statistically significant. For the final analysis, therefore, the two groups were pooled, and the results of the summational analysis rotated by the group factor method, using the procedure recommended in [29][2].

[1] The experiments were originally started at the London Day Training College (subsequently known as the Institute of Education) with small groups of children and somewhat elaborate apparatus, and after the war were resumed first at University College and then on a more extensive scale at various schools of different types. At every stage they were carried out in consultation or collaboration with other investigators, more particularly Mr. R. B. Lewis and Miss Margaret Howard, who, it is hoped, will contribute some of the later papers. To Professor Burt, Sir Percy Nunn, and Professor Hamley we owe a deep debt of gratitude for their help and advice.

[2] The rotation was carried out quite automatically by two co-workers who knew nothing of the hypotheses to be tested. For help in this respect we are indebted to Miss R. L. Lucas and Mr. W. G. Morton.

Individual Tests. Although our interest was largely in modern scientific reasoning, we avoided problems or apparatus of a conspicuously scientific type. In this country, owing to the advocacy of Professor H. E. Armstrong [1][1] and Sir Percy Nunn, science teachers became greatly impressed with the value of the ' heuristic method ' : this consisted in confronting the pupil with simple, practical, and intelligible problems in physics or chemistry, and encouraging him to hammer out for himself a suitable procedure and reach his own conclusions. And in the early days of child study, many British schoolmasters kept systematic observations on the methods adopted by older children (chiefly boys) in their attempts to solve such problems in the school laboratory or classroom. It is evident, however, that experiments on the stock questions of elementary science, especially when laboratory apparatus has to be used, are greatly affected by the experience which different children have already enjoyed—by the knowledge they have picked up at home, by the information gathered from children's encyclopaedias and periodicals, or from their training as boy scouts, and by their previous familiarity with scientific toys, meccano sets, children's conjuring sets, and the like. Nor are the tasks themselves of very great interest to the ordinary child.

We tried to base our own problems on children's well-known games or pastimes. The apparatus used consisted of objects and appliances with which nearly every child would be familiar : coloured beads, papers, blocks, and dress materials, lamps, mirrors, balls, a see-saw, pulleys made of cotton-reels, a home-made pair of scales, a home-made pendulum, a home-made spirit level, and so on[2]. With older children we made free use of verbal problems ; and in studying compound forms of reasoning we adopted the familiar device of converting the commoner logical patterns into concrete form by camouflaging them as spy or detective stories where the conclusion would follow as a necessary inference from the premisses or ' clues '. Examples of the more practical and less familiar type of problem (often adapted from those of Burt, Hamley, and Isaacs) will be found in the appendix. We made no attempt to grade the several problems in order of difficulty. Instead, we relied, for our statistical data, on a grading of the commoner solutions.

[1] Older teachers will recall Armstrong's various stories—e.g. ' The Monkey who would not be killed ', an ingenious creature whom its owner flung into a pond with a stone tied to its neck; in water the stone lost some of its effective weight, so the monkey held its breath and simply walked ashore along the bottom of the lake: " was that really possible ? ", asks the teacher, and the pupils are told to discover the answer for themselves. Nunn's own lectures and papers on the teaching of science (many unfortunately still unpublished) contain many valuable first-hand observations on the reasoning of children when encouraged to investigate for themselves the laws of falling bodies by Galileo's device of an inclined plane, and the elementary laws of optics, hydrostatics, magnetism, and chemical composition. See also ' Science Teaching in School ', *ap. Education in the Nineteenth Century*, by C. W. Kimmins, head of the L.C.C. school inspectorate, who did so much to facilitate the researches of psychologists during the first quarter of the century.

[2] Home-made because manufactured instruments either do not exhibit the underlying principles very clearly or suggest a scholastic or scientific background, whereas we wanted to study the child's understanding of problems encountered in his own out-of-school life.

Subjects. For the factorial studies we applied the group tests to 300 children, aged between 10 and 11, drawn from different schools, and so selected that their I.Q.'s fall within the narrow range of 95 to 105. Actually only 247 out of the total number were able to take all the tests ; and it is on these that the tables below are based. For the more intensive studies with individual tests, we selected first children showing exceptionally high or exceptionally low factor measurements in each of the factors, and secondly median children (i.e. children whose mental age corresponded to their chronological age) drawn from every age group from 5 to 15. With the Mirror and the Ball-and-Wall tests we also tested what were intended to be random (though small) samples of every age from 7 to 13. In my own experiments the children tested were mainly girls ; in those of Mr. Lewis, they were chiefly boys.

III. Results

1. *The Components of the Reasoning Process.* The first question to be asked about any definitely distinguishable type of intellectual process relates to its essential nature. Does it involve a specific ability over and above so-called general intelligence ? If so, is this a simple ' primary ability ', or does it involve several distinct components ? In the case of reasoning Burt has put forward the hypothesis that efficiency in reasoning depends both on general intelligence and on specific factors [28] ; Spearman and Wynn Jones maintain that " the more recent results are unanimous in denying any specific factors that can be designated by such terms as ' reasoning ', ' deduction ', ' induction ', or even ' evidential relation ' "[1].

In our own experiments the effects of general intelligence were largely eliminated by the way the children were selected. Nevertheless, with both boys and girls most of the tests of reasoning still showed small correlations with the test of intelligence. This test was accordingly used to partial out, so far as possible, any residual influence of intelligence remaining in the intercorrelations before they were combined. The coefficients so obtained were first subjected to an analysis by Burt's method of ' simple summation '. Four significant factors were obtained—a general factor, which many no doubt would identify with a factor of ' general reasoning ability ', and three bipolar factors, suggesting by their sign patterns the presence of six or seven group factors. These were then rotated to obtain group and subdivided factors. The saturation coefficients for the rotated factors are shown in Table I.

Judged by the usual formula for testing the significance of factor saturations [32, p. 128], it may be said that broadly speaking any saturation over 0·13 is statistically significant. To determine the nature of the various

[1] Spearman, C. and Jones, L. W., *Human Ability*, 1950, p. 92. ' Evidential relation ' is Spearman's name for what Burt commonly calls ' implication ' and other writers ' entailment '. Spearman here is not denying the ability to perceive or ' cognize ' this particular relation, but only the theory that " sets up such an operation on a pedestal as appertaining to some higher power of reasoning " (*Nature of Intelligence*, p. 72).

TABLE I. TESTS OF REASONING AND ALLIED PROCESSES

Saturations for group factors obtained by rotation

Test	I	II	III	IV	V	VI	VII	VIII
1. Concept Formation (V)	0·139	0·422	0·056	0·503	0·151	0·222	0·315	−0·038
2. Concept Formation (N)	0·111	0·565	0·034	0·486	0·197	0·165	0·181	0·206
3. Generalization (N)	0·127	0·316	0·102	0·611	0·243	0·241	0·157	0·321
4. Abstraction (V)	0·143	0·131	0·437	0·542	0·314	−0·014	0·363	0·026
5. Judgment (V)	0·216	0·152	0·181	0·178	0·175	0·082	0·278	0·078
6. Judgment (N)	0·160	0·243	0·093	0·214	0·218	0·117	0·094	0·215
7. Simple Relations (V)	0·682	−0·068	0·108	0·072	−0·122	0·098	0·186	0·056
8. Simple Relations (N)	0·723	−0·080	0·075	0·187	0·066	0·153	0·045	0·189
9. Analogies (V)	0·642	0·075	0·409	0·051	0·141	0·066	0·171	−0·034
10. Analogies (N)	0·665	0·091	0·321	0·219	0·198	0·186	0·034	0·095
11. Matrix (N)	0·434	0·617	0·020	0·502	0·242	0·061	0·042	0·341
12. Easy Syllogisms (V)	0·381	0·084	0·626	0·109	−0·187	0·028	0·133	0·022
13. Harder Syllogisms (V)	0·472	0·514	0·513	0·062	0·461	−0·014	0·166	0·078
14. Series: 3 Terms (V)	0·515	0·245	0·212	0·085	0·059	0·032	0·223	0·103
15. Series: 3 Terms (N)	0·428	0·351	0·063	0·191	−0·113	0·393	0·051	0·159
16. Series: 4 Terms (V)	0·396	0·083	0·265	0·122	0·224	0·320	0·234	0·065
17. Series: 4 Terms (N)	0·383	0·172	0·102	0·213	0·186	0·419	0·096	0·251
18. Hypothetical Deductions (V)	0·414	0·393	0·524	0·148	0·373	0·041	0·131	0·077
19. Disjunctive Deductions (V)	0·328	0·463	0·558	0·085	0·409	0·803	0·159	0·134
20. Misc. Compound Deductions (V)	0·274	0·550	0·614	0·027	0·513	−0·021	0·182	0·069
21. Easy Inductions (V)	0·120	0·612	0·313	0·410	−0·055	0·096	0·210	0·111
22. Easy Inductions (N)	0·235	0·743	0·183	0·361	0·124	0·162	0·073	0·274
23. Harder Inductions (V)	0·316	0·586	0·305	0·392	0·395	0·015	0·237	0·072
24. Harder Inductions (N)	0·217	0·682	0·227	0·318	0·432	0·202	0·112	0·297
25. Span of Attention (N)	0·121	0·312	−0·029	0·150	0·483	0·134	0·018	0·103
26. Free Association (V)	−0·137	0·106	0·124	0·096	0·146	0·497	0·184	0·064
27. Fluency (V)	−0·162	0·131	−0·046	0·164	−0·061	0·563	0·217	−0·049
Sum of Squares	3·747	4·120	2·511	2·341	1·966	1·333	0·871	0·687
Contribution to Variance (per cent)	13·9	15·3	9·3	8·7	7·3	4·9	3·2	2·5

factors, children having extremely high or extremely low factor measurements for each of the factors obtained were then subjected to a series of individual tests ; and an attempt was made to get each child to describe the methods he used and to furnish introspections regarding visualization, verbalization, and the like.

Except for two supplementary tests (which are not tests of reasoning) Factor I has positive saturations throughout, though several are not fully significant. A study of those children having high and low factor measurements indicates a factor for the conscious apprehension (explicit or implicit) of relations as such, as distinct from the apprehension of concrete percepts or the eduction of concepts by unconscious associations.

Factor II yields the highest contribution to the total variance. It may be roughly described as a factor for the combination of relations in accordance with the principles of logical validity. Children with the highest factor measurements were distinguished by their capacity for rearranging a set of data presented in irregular or incomplete form, and organizing them into logically coherent relational systems. It has many features in common with Maxwell Garnett's factor of ' cleverness '.

This factor is apparently subdivided into two narrower group factors—Factors III and IV. As Burt pointed out some 40 years ago [4, p. 104], " nearly all the relational tests, and nearly all the methods of attacking relational problems, seem divisible into two contrasted groups " (indicated in his own experiments by " a marked and recurrent bipolar factor ") " so that the comprehension of a relational complex can apparently be reached in two different ways, explicit or implicit " [28, p. 192]. This seems to be confirmed by our study of the children with the highest factor measurements for each of the two group factors. The characteristic method of the one batch is analytic or explicit reasoning, proceeding by a succession of logical steps ; of the others, an implicit or synthetic apperception of what many psychologists would call Gestalten or configurations, resulting from a kind of comprehensive insight which integrates the constituent aspects almost simultaneously and so enables them to jump to the right conclusion[1].

[1] Tachistoscopic tests suggest that those with a high factor measurement for Factor III commonly belong to the so-called ' fixating ' type; they generally begin by dissecting the complex whole presented to them into successive parts, and so apprehend the connecting relations explicitly: they then go on, if the problem requires it, to apply those relations in a systematic fashion to elicit a suitable solution. As a rule, they are more successful with problems requiring deductive inferences (particularly those soluble by formal syllogistic reasoning) than with problems requiring an inductive inference. Children with high factor measurements for Factor IV show the characteristics of what is called (not very appropriately) the ' fluctuating ' type of attention [2]. They seem to rely more upon " a complex synthetic activity comparable to the activity often popularly described as intuition—the activity whereby we implicitly comprehend the essential character of a whole, without analysing it into its component parts, or distinctly formulating their relations " [4, p. 104]. They are especially successful with test problems involving generalization—the formation of general concepts or the perception of general laws or rules, as well as with what are popularly termed inductive problems. And, as the test-saturations imply, they are better at problems in spatial or pictorial, i.e., visual form, whereas the former are better at verbal. Much the same contrasts may be noted with adults: thus among University students the ' analytic ' type seems to include most of those who possess high scientific ability; the ' intuitive ' type most of those who possess a marked gift for aesthetic appreciation [28].

Burt and Corkhill apparently found a similar subdivision in the factor for the apprehension of relations (our Factor I). Their two group factors they regard as factors for the explicit and the implicit apprehension of relations as such. In our own experiments the preliminary summational analysis showed some slight indication of a corresponding bipolar factor ; but the saturations were not statistically significant. And, after rotation, it would seem that the two factors noted by Burt and Corkhill become absorbed in the two group factors for what we have called the analytic and intuitive types respectively : in fact, it seems doubtful whether they really form distinct and separate pairs.

Of the remaining factors, Factor V appears at first sight to be merely a difficulty factor. But its largest saturation is for the test of span of attention. Hence it may perhaps be due to the fact that the methods adopted for eliminating differences in general intelligence were not completely effective.

Factor VI is evidently a factor of productive association. And the last two are factors for content, Factor VII being the familiar verbal factor, and Factor VIII a visuo-spatial factor or rather (as the introspective comments suggest) a factor for ability to visualize. There were minor indications suggesting the presence of small factors for the apprehension of different kinds of relations : these would also have to be classed as factors for content rather than as formal factors. But none was statistically significant.

Schemata. One problem still remains. If, as Burt maintains and our figures tend to suggest, the factor for combining relations may be plausibly regarded as the ' general factor ', or at least the most important factor, in the processes of reasoning, how precisely does the successful reasoner achieve the right type of combination ? There seem to be two alternative hypotheses—that of Spearman and that of Burt.

Burt points out, to begin with, " that not any kind of coherent combination will suffice: an imaginative child is quite capable of combining all the data accessible to him into a plausible interpretative pattern that has no logical validity whatever. To reach a correct solution, the child has therefore to select those types of pattern that satisfy the laws of valid inference ". The child is of course quite unfamiliar with the laws of logical syntax just as he is usually unfamiliar with those of grammatical syntax. The latter he " picks up by ' incidental learning ' in the form of implicitly apprehended *schemata*—blank frames or moulds that are filled in by the concrete data and what the data suggest, and so provide a plan of operation[1]; the same holds true of the processes of reasoning: he learns to argue logically in the same way that he learns to speak grammatically "[2].

[1] " The grouping of parts of speech into sentences, in the appropriate order and with the appropriate inflections, involves a general apprehension of schematic form, guiding and controlling the serial succession of ideas and words " (Stout, *Analytic Psychology*, 1909, II, p. 37). The first explicitly to recognize this process seem to have been the grammarians themselves: cf. M. H. Paul's discussion of ' word-combination by analogy ' (*Principles of the History of Language*, English translation 1890, chap. V.).

[2] The concept and the name ' schemata ' were suggested by Stout's discussion of the part played by ' noetic synthesis ' in apperceptive processes. " So far as an implicit idea of perception of a whole determines the successive emergence of the parts, we may apply to it the term ' schematic apprehension ': . . . the ' implicit idea of the whole ' operates as a *schema* to be filled in—a general awareness of the form of combination apart from the actual elements to be combined, a form which receives progressive specification as the train of thought proceeds " (*loc. cit.*, I, pp. 95f.; cf. also

Spearman, on the other hand, rejects the whole notion of specific factors for reasoning. Reasoning, he holds, is merely one of the many manifestations of the general factor of ' intelligence '. Hence he will have nothing whatever to do with " what are variously named ' structures ', ' integrated wholes ', ' modes of apperception ', ' forms of noetic synthesis ': here as elsewhere such concepts have only wrought confusion ". Instead, he maintains, " all the relevant facts can be adequately accounted for by the two noegenetic principles—the eduction of relations and of correlates " (*loc. cit., sup.*).

Today few psychologists accept this latter explanation. However, the only satisfactory way to decide between the two hypotheses is (in our view) to study the incidental hints and indications furnished by the children's own introspections and by first-hand observation of the way each child goes to work. But this is a supplementary issue which must be postponed to a later paper. Here it will be sufficient to say that our own results seem amply to confirm the alternative explanation of ' determinative schemata '.

2. *The Hypothesis of Developmental Stages.* We now turn to consider the way in which the reasoning process appears to develop. Here also two quite opposite views have been put forward. Most psychologists of the biological school—Spencer, Stanley Hall, Claparède, and Piaget—and the vast majority of educationists and teachers, hold that mental growth proceeds by stages, and that the ability to reason marks the last stage of all, appearing only with the approach of puberty, i.e. about the age of 12 or shortly after[1]. Burt and Isaacs,

Selz [7]). The nature of the schemata is best expressed by the use of a symbolic notation which presents us with an abstract model, as it were, of the formal structure involved. The term ' schema ' itself goes back to Kant (*Kritik der reinen Vernunft*, 2te Ausg., Thl. II, pp. 139f.). (I am indebted for these and the previous references to Professor Burt.)

[1] Stanley Hall, for example, argues at length that " reasoning is the capacity which evolves last in the human race and consequently appears last in the development of the individual ". Mental growth he divides into a number of successive stages, corresponding to those of mental evolution—reflexes, sensori-motor control, habit and memory, and " lastly comes the stage of rational thought ". It follows that " the child cannot reason according to adult standards until fourteen—the age at which Aristotle would begin the education of reason " (*Psychology of Adolescence*, 1905, pp. 111f. cf. *Aspects of Child Life and Education*, 1907, pp. 1f.). Hall and most of those who adopt similar views hold that such changes are primarily due to the " maturation of the brain, particularly no doubt a multiplication of cells in the association areas ". In point of fact, as recent histological studies show, the actual multiplication of *cells* is over before the child is born; and, though the proliferation of *dendrites* in the frontal and other association areas continues till the age of four, it appears to have ceased entirely long before the age of seven.

Piaget also declares that " formal thought reaches fruition during adolescence ". He distinguishes somewhat similar stages in the child's intellectual development: (i) the sensori-motor stage; (ii) habits learnt by trial and error; (iii) preconceptual thought; (iv) intuitive thought (4–7 yrs.); (v) concrete operations (7–11 yrs.); and then (vi) " beginning at 11–12 years, there is a period of new structuring, which is not simply the culmination point of the 7–11 stage, but bound up with an entirely new set of operations—a formal mode of thought, leading to (vii) another level of equilibrium at about 14–15 years ".

In explaining this final stage he says " a neat example comes from one of Burt's tests: ' Edith is fairer than Susan; Edith is darker than Lily: who is the darkest of the three ? ' This is rarely solved before the age of 12. Till then we find reasoning such as the following: ' Edith and Susan are fair; Edith and Lily are dark; hence Lily is darkest ' ". And he concludes that " formal logic is still inaccessible to the child " (*enfant*, which in France usually means child under 12: *The Psychology of Intelligence*, 1950, pp. 148f.). Here, however, Piaget seems to have misunderstood Burt's account of the results obtained with this test. To begin with, it is called ' *test* 12 ', not a ' test for *age* 12 ': it is answered by 46 per cent of the eight-year-olds and 78 per cent of the 10-year-olds. Possibly, as Dr. Isaacs points out, the I.Q.'s of Piaget's children were lower. But the question actually put in the test is, not ' Who is the darkest of the three ? ' but ' Who is darker —Susan or Lily ? ' Burt goes on to describe how, with the eight-year-old groups, the last premiss

on the other hand, reject the whole theory of development as proceeding by a succession of qualitatively different stages, and with it " the widespread notion that during these earlier years (7 to 11) the child cannot reason " [11, pp. 255–7, 265–8]. In their view (as we have seen) " all the mental mechanisms essential to reasoning are present before the child leaves the infants' department, i.e. by the mental age of seven, if not somewhat before ", and subsequent development consists chiefly in " an increase in the complexity of the processes which the child is able to perform, that is to say, in the number, variety, originality, and compactness of the relations which his mind can apprehend and integrate into a coherent whole " [5, p. 127].

Results of Group Tests. The results of our own tests clearly show, not only that the problems we have used may be arranged in order of increasing difficulty, but also that this order is fairly stable. Taking first the group tests, and ranking them in order of difficulty for different individuals, we find that in nearly every case the ' interclass ' correlation is over 0·80, except where special knowledge, experience, or interest has affected the child's performance. And, if we now rearrange the problems in the average order of difficulty, and calculate the proportion of children successfully solving each problem in each successive age-group, then it is clear beyond question that there is a steady improvement in reasoning ability throughout the whole school period. Moreover, if we classify the various problems according to the threefold classification of logical processes adopted by many modern logicians (e.g. [17], p. 31) into the calculus of ' classes ', ' relations ', and ' statements ' (or ' propositions ') respectively and next divide the list of tests into three approximately equal sections—' easy ', ' moderately easy ', and ' difficult ', we then discover that 76 per cent of the first type fall into the easy section and 52 per cent of the third group into the ' difficult ' section. Thus, it is tempting to infer that the ' calculus of classes ' develops earlier than the ' calculus of relations ' (as in problems involving serial arrangements), and the ' calculus of relations ' than ' the calculus of statements '. There is, however, much overlapping, particularly between the two last. And Burt's conclusion is more cautious, namely, that " in general, the calculus of classes is easier than the calculus of propositions ", and accordingly " the tendency to ' propositionize ' (to borrow a word from

' still ringing in the child's memory ' tends to dominate the rest, so that Edith and Lily are regarded as belonging to the dark group; consequently the child selects Lily as the dark one in contrast to Susan, who is associated with the word ' fair '. (With Piaget's form of the question, we find, the more usual mistake is for the child to say, not Lily, but Edith, since she is the darkest of the dark pair.) However, as Burt observes, " the problem has only to be rephrased ' Lily is fairer than Edith; Edith is fairer than Susan: who is the fairest ? ' for the question to be answered by 72 per cent at the age of 8, instead of by only 46 per cent ". And he adds that the children's " introspective comments " indicate that " nearly every child who answers correctly [with the original form of the test] actually converts the disparate premiss, so that the problem becomes a problem in an ordered three-term series " [5, p. 126]. Thus, the difficulty of the original version of the test is due, not to the nature of the formal reasoning, but to the fact that it involves three steps instead of one. (On the general questions raised by this type of reasoning, see the instructive study by Ian Hunter ' The Solving of Three-Term Series Problems ', *Brit. J. Psychol.*, XLVIII, 1957, pp. 286–298. For a more detailed examination of Piaget's hypothesis see Dr. Isaacs ' discussion [*op. cit. sup.*, pp. 77f.].)

Stout) seems far more characteristic of the older and more intelligent of our reasoners ". " The reason," he suggests, " is that the former involves thinking in terms of what may be called (perhaps in a slightly wider sense than usual) the ' object language ' and the latter thinking in terms of the ' meta-language ', that is to say, thinking which proceeds by making statements about statements—a procedure which is obviously more complex. "

To explain the difference he describes two typical examples. " A boy of $7\frac{8}{12}$ years (I.Q. 96) was asked to take from two open boxes (which he called ' lids ')—a round box and a square one—a number of flat blocks, some circular like draughts, and others square or nearly so like dominoes, and to solve certain problems. When the game was over, he was asked to put the blocks back in their boxes, and picked up the square box. I handed him two or three round blocks. After a moment's thought he pushed them back, saying, ' No, the blocks in the square lid are all square blocks. Your blocks aren't square. So they don't go in the square lid '. Here the child's reasoning is in terms of classes. It is a straightforward syllogism in the Mood *Cesare* of the Second Figure. In the traditional symbolism its formal structure would be exhibited as follows[1]:

$$\begin{array}{l} S\ a\ M \\ P\ e\ M \\ \hline \therefore\ S\ e\ P \end{array}$$

" Another boy, aged $7\frac{5}{12}$ (I.Q. 108), was arguing with a play-fellow (aged 5) about some frogs they saw leaping about in the grass near the pond. ' Why do you keep calling them " fish " ? ' he said; ' my Daddy says fish can't live out of water. All these frogs can. So it's wrong to call them fish.' Here the reasoning is in terms of propositions. Expressed in the traditional symbols, the scheme would be the same as before. But that would conceal the fact that a new and more powerful type of reasoning is being employed—the ' calculus of propositions ' instead of the ' calculus of classes '.

" To express the difference therefore let us borrow the two new types of symbolism introduced by Russell and Whitehead and adopted (with various modifications) by most logicians[2]. In the notation for the ' calculus of classes ', the first argument may be expressed as follows:

$$\begin{array}{rl} & S \subset M \\ P \subset \overline{M} & \therefore\ M \subset \overline{P} \\ \hline & \therefore\ S \subset \overline{P} \end{array}$$

[1] Cf. Welton, *Manual of Logic*, I, p. 320, or S. Stebbing, *A Modern Elementary Logic*, pp. 66f. For those who are unfamiliar with the traditional notation it should be explained that S and P denote Subject and Predicate of the conclusion, M the Middle term, a " All . . . is . . . ", e " No . . . is . . . ". The names for the different types of syllogism, e.g., *CeSaRe* for the scheme *e. .a. .e* (the consonants have no significance), were introduced by the 17th century logicians (H. Aldrich, *Artis Logicae Rudimenta*).

[2] [3] cf. [25]. Burt, however, retains several older conventions—e.g., the bar (for negation), brackets (to indicate the order in which symbols are to be combined), and the displaying of formulae (instead of Frege's assertion-sign) to show that propositions are asserted and not merely considered. Thus the first line of the final syllogistic scheme (set out on p. 150) may be read as follows: " It is asserted as true that for every x, ' x is an m (a frog) ' implies ' x is a p (a creature that can live on land) ' ". Note that the conversion of the single letters S, M, and P into the functional notation serves to indicate the internal structure of those expressions, which in many instances would otherwise be ambiguous and lead to fallacies.

where $\subset$ denotes the transitive relation of ' class inclusion ', and the bar the negative. With the notation for propositions the second can be expressed as follows:

$$\begin{array}{ll} & (x)\ [s(x) \supset m(x)] \\ (x)[p(x) \supset \overline{m(x)}] & \therefore\ (x)\ [m(x) \supset \overline{p(x)}] \\ \hline & \therefore\ (x)\ [s(x) \supset \overline{p(x)}] \end{array}$$

where $\supset$ denotes the transitive relation of ' implication '. Here I have reversed the order traditionally adopted and have placed the minor premiss before the major, so that the transitivity is obvious. This in fact is the order in which the young child reasons in such cases when he reasons successfully."

This broad difference between the reasoning of younger and older children is amply confirmed by the explanatory comments given by the children we ourselves tested. Nevertheless, when occasion requires (e.g. in the case of a dispute between two or more children) even quite a young child will drop into the ' meta-language '. And the difficulty of the latter mode of reasoning springs, not from some qualitative difference which the younger child is unable to master, but merely from its greater complexity [1].

Results of Individual Tests. Most of those who claim to find radical changes at or about certain ages speak as though they were referring to *chronological* ages, which apparently they suppose to be characterized by the maturation of certain physical structures in the brain or cortex. The evidence already cited sufficiently proves that, so far as individuals are concerned, this view is quite untenable. Moreover, the assessments obtained for general intelligence show that, within each single age-group, the intelligence of the individual children may vary as much as 8 to 10 mental years—a variation which is obscured when we consider only the typical performance at each age; and the correlations with intelligence show that success in reasoning tests depends largely upon the degree of intelligence possessed by each individual. In view of these well-known results one or two recent writers have argued that it is still conceivable that qualitative changes may occur at certain *mental* ages.

To deal with the question in this modified form, we must now turn to the results obtained with children tested individually. In order that the comparisons may be based on mental ages rather than chronological, we selected from each of our age-groups (which ranged, it will be remembered, from 7 to 13) 10 children (5 of either sex) whose I.Q.'s varied less than 5 points from 100, and whose mental ages therefore were roughly identical with their chronological ages. Not all the individual tests described above could be applied to every child. Two, however—the ' mirror ' and the ' ball-and-wall ' test—were taken by all the 70 children. Instead of grading the tests according to their relative

It will also be noted that the schematization adopted for the mood *CeSaRe* inverts the second premiss (in accordance with the rule for ' contraposition '), and so reduces the mode of reasoning to the simplest and commonest type of elementary syllogism, namely, *BaRBaRa*, the First Mood of the First Figure; and the inferential process is thus greatly clarified. Once again this is the procedure adopted, explicitly or implicitly, by most children when they solve such problems successfully.

[1] No doubt, the maturation of intelligence with increasing age may be one factor in facilitating the change; but the main influence is the child's free intercourse with other children, particularly when the latter vary in age and understanding.

difficulty, the children's answers or procedures were graded according to their relative merits. Their final solutions, i.e. the best approach each child could make towards a rational solution, were classified into a series of eight groups, each more satisfactory than the last. The groups or grades may best be indicated by selecting the most typical answer or performance in each case. For the two tests just mentioned they were as follows :

i. Complete failure.

ii. " You have to trust to luck: sometimes you get it (the mirror) to shine on the boy's face; sometimes it doesn't." " Sometimes the ball goes back to the boy; sometimes it doesn't."

iii. " You hold the bit of looking glass so that it turns towards the boy." (*Experimenter*: " So that he could see his own face in it, if he was near enough ? " " Yes.") " You aim the ball at the point on the wall opposite where the boy is standing." (*Experimenter*: " So that it goes straight back to him, like this ? " " Yes.")

iv. " You keep moving the bit of looking glass about, and in the end you find the right way to hold it." " You keep aiming the ball in different directions until you find the right one."

v. " You don't hold the looking glass pointing towards the boy, but somewhere between him and the light." " You aim the ball at a place on the wall somewhere between you and the boy."

vi. " You have to hold the glass so that it looks towards a point half-way between the light and the boy." " You have to aim the ball at a place on the wall half-way (midway) between you and the boy." (Or any formulation implying that there is a definite rule, though the child fails to express it correctly and his actual attempts show that he has not discovered the right rule.)

vii. " The light shines from the lamp on to the mirror, and then turns back again like a ball bouncing off the glass." " You have to send the ball sideways to the wall, and then it will bounce away to the boy over there." (Or any formulation implying that there is a definite rule, though the child cannot formulate it explicitly, provided his performances indicate that he has got the right principle.)

viii. The boy is able to draw the path of the direct and the reflected ray (or the movement of the ball to the wall and from the wall) and to indicate in his own fashion that the angle of incidence would be equal to the angle of reflection (or rebound), e.g. by bisecting a line drawn from the incident ray to the reflected ray, parallel to the mirror. And his performances show that he can apply the principle when his position, or that of the light or the boy, is changed. (Only one of our testees used the word ' angle ', but three said " the slope had to be the same ".)

The number in each age group giving each type of answer is shown in Table II.

Were we to trust to the results from the Mirror test alone, we might be inclined to argue that there were fairly well-marked changes at certain mental ages in the types of solutions suggested : no one, for example, hits on the idea that there may be a definite rule governing the reflection of light before the mental age of 10 ; and not until the age of 12 or 13 is any child able to reach a working generalization as to what the rule is. However, on turning to the results of the Ball-and-Wall test, we see that similar types of solution are reached much earlier. One of the 8-year-old children already has some notion that there is a guiding rule ; and one of the 10-year-olds has grasped it accurately enough for all practical purposes. Evidently the age at which the

changes in type of solution appear is related, not so much to formal differences in logical reasoning, but rather to the material nature of the problem. The path of the ball can actually be seen ; that of a ray of light cannot. Moreover,

TABLE II. MIRROR (M) AND BALL-AND-WALL (B) TESTS

Number of answers of each type at each successive age

Age	7		8		9		10		11		12		13	
Test	M	B	M	B	M	B	M	B	M	B	M	B	M	B
Type of Response														
i	8	5	5	3	4	1	2	1	1	0	0	0	0	0
ii	1	2	4	3	3	2	2	1	1	0	0	0	0	0
iii	1	2	1	1	1	3	3	0	0	1	1	0	1	0
iv	0	1	0	1	2	1	1	2	3	1	1	1	1	0
v	0	0	0	1	0	1	1	3	3	4	3	2	1	1
vi	0	0	0	1	0	2	1	2	2	2	3	3	3	2
vii	0	0	0	0	0	0	1	0	0	1	2	3	3	4
viii	0	0	0	0	0	0	0	0	0	1	0	1	1	3

both boys and girls are far more familiar with the behaviour of balls than of rays of light, and (in an implicit fashion) with the laws governing the motion of solid objects than with the laws of optics. The younger boys, it may be added, are distinctly better with the Ball-and-Wall problem than the girls ; while the girls seem slightly better acquainted with the properties of mirrors and of light. Two girls had even realized, from seeing rays of sunlight in a dusty shuttered room, that light travels in straight lines.

The method which the child adopts in trying to discover a solution seems to depend quite as much on temperament as on the development of reasoning ability. The boy who tries to think out a reason for the phenomena, and to discover from actual trial whether his own explanation is right, is an inquisitive and not necessarily a logical child. And it is the lively and slightly impudent youngster who starts moving the lamp about and changing his place. Even the substitution of formalized reasoning for concrete reasoning (or, as Burt would say[1], of the ' meta-language ' for the ' object language '—a change which

[1] As he observes: " The first step towards the meta-language is taken when the child substitutes a hypothetical proposition for a universal categorical in the major premiss of a syllogism. The change in wording usually implies a change from an inquiring attitude to a controversial." When a small boy, telling you about a bat he has caught in his toothglass, says " Birds have feathers; this hadn't; so it wasn't a bird ", he is just reporting his own private thoughts. When he says, " If it was a bird, it would have feathers ", he is arguing with an actual or an imagined companion, who has made a false statement. And when he says, " The way it hangs upside down shows it can't be a bird; that's silly ", he is explicitly using the meta-language. He is virtually stating, " ' This creature hangs upside down ' implies ' This creature can't be a bird ' ". Here the *quotation marks* indicate that the reasoner's statements are primarily about propositions, and no longer about the objects (though of course the propositions themselves make statements about the objects). No doubt statements of the latter type need greater intelligence, if only because of their increased complexity; ' Birds have feathers ' is a single clause, containing three words only; to change this to ' If it was a bird, it would have feathers ' means organizing nine words and two clauses into a

Piaget regards as a new and distinctive step achieved only towards adolescence) seems due rather to the child's emotional attitude than to his cognitive capacity : it is the aggressive, argumentative child who tends to express his reasoning in the form " It isn't true that (so-and-so) because, if it were, then you couldn't say (so-and-so) ". Much depends on the social context of the inquiry : a young child working by himself is likely to talk or think chiefly in the object language ; but when two or more children start wrangling together, then each changes to the meta-language to justify his own conclusions[1].

In any case, it would manifestly be wrong to describe the results recorded above as characteristic of mental ages as such. Thus a youngster of 7 with a mental age of 10 (I.Q. therefore about 140) will produce much better theoretical solutions than those assigned to the mental age of 10 in the table, where the mental ages cited are those of children of *average* I.Q. ; on the other hand, the older child is likely to be more facile in the practical manipulations. As a result, in my own experiments at least, I.Q.'s calculated from the scale of reasoning tests vary far more widely about the mean than the I.Q.'s calculated from an ordinary test of intelligence. Children with an average I.Q. of 125 on the London revision of the Terman–Binet have I.Q.'s averaging 143 with Burt's scale of reasoning tests : their ' reasoning quotient ' (mental age for reasoning ÷ mental age for intelligence) amounts to nearly 115. And generally the development of reasoning in the sharp precocious child is far more rapid and follows quite different lines, as compared with that of the ordinary child. Here, therefore, is a further reason why, even if we interpret the ' ages ' mentioned by the supporters of the ' stage-hypothesis ' to mean ' mental ages ', their conclusions cannot be accepted.

3. *The Development of Specific Relations and Schemata.* Burt holds that with the average child certain types of relation are apprehended[2], and certain

single meaningful assertion—a far harder task. And it is only the more experienced that are able to distinguish between statements expressing assumptions, hypotheses, or inferred conclusions, and statements reporting direct observations.

[1] This is probably what Dr. Isaacs has in mind when she says that " sustained conversations between one child and one adult (the method adopted in Piaget's experiments) do not suffice to reveal the child's full capacity for rational argumentation ". Personally, however, we find that, when the child accepts the supervisor as an associate on equal terms, then the questions and views of the latter *do* lead the child to think explicitly about his thought-operations and about the statements that express them, as is shown by his frequent use of ' if ', ' because ', ' you can't say ', ' that shows ', etc.

[2] The relation of ' difference ' and of ' subject and attribute ' well before the age of 7; of ' part and whole ', and ' similarity ' and ' degree ', by about 8 or 8½; of ' space ', by about 9; of ' time ' later still; and ' cause and effect ' last of all. He holds that all except the last are apprehended directly and have an objective existence. On the other hand, in his view, the relation of cause and effect is not an objective relation, though the underlying relation of implication is; and the use of the former by the growing child varies according to the extent with which he becomes familiarized with the concept by references to it at home or by instruction at school. Susan Isaacs, on the other hand, seems to hold that the apprehension and use of the cause-and-effect relation develops quite early.

types of argument used[1], at a later age than others [5]. In my own results I have found no clear differences in difficulty between the commoner types of relations, but only between the different types of argument. And in neither case (so far as I can see) are there any which could not be explained by assuming that simpler and more concrete conceptions (especially visual or visuo-motor) are grasped more readily than more complex or more abstract conceptions.

The chief characteristic therefore distinguishing the thought-processes of the better and the more mature reasoners from those whose thinking tends to be largely illogical (i.e. chiefly imaginative or mechanical reproduction) would seem to be the facility with which they can work explicitly with highly complicated *schemata*. To determine the extent to which they succeed, however, a different type of problem is needed—tasks in which a fairly elaborate train of thought is required. Owing to the excessive time needed for individual testing of this kind, our own studies are so far rather meagre. Here, therefore, I must content myself with briefly illustrating the general method and its results by one of the shorter and more effective of such tests—the Code test described in the Appendix.

With this a quantitative mark for the child's solution can be at once obtained by merely counting up the number of correct steps that it contains. But that of itself throws little light on the nature of the schema that the child is using. Accordingly he must be persuaded to formulate aloud the various stages in his train of thought as he goes along. If these are jotted down at the time and then further elucidated by tactful questioning, the underlying plan of operation that, half unconsciously, guides the child's reflections can be elicited without much difficulty.

Let the reader work out the solution for himself, and make a diagram of his various steps; he will find that he is using a schema which includes several different types of inference which together form, not a simple serial train of thought, but rather a branching argument. The first step of all is the most elusive. The reasoner must begin by deciding how to attack the problem: this is perhaps the most important stage of all, and one that can seldom be traced to any explicit kind of argument.

The less able pupils commonly start by trial and error. The more efficient hit on a hopeful opening by semi-intuitive insight. All such methods the teacher would probably describe as ' guessing '. But, unless the problem is quite beyond the child, such guessing is seldom wholly blind or random. Nearly always some intelligible motive, usually unconscious, prompts the child to choose his initial line of attack. Frequently it is suggested, not by the special features of the problem itself, but by his own past experiences. And the method adopted at times appears indicative of the intellectual type to which the child belongs—intuitive or analytic: as a rule, however, a mixed type seems the most efficient.

Among those who succeed the commonest procedure consists in an application of the ' hypothetico-disjunctive ' schema. With the present problem, by a kind of rapid intuition (a ' schematic apprehension ', as it were, of its rarer features) the child sees almost at a glance that the easiest words to start on will be either the short words with double letters or the long words. In the list of clues there are only two words belonging to each type,

[1] Syllogistic by age 7; disjunctive by age 8; hypothetical soon after; the *modus ponens* ($[(a \supset b) \,.\, a] \supset b$) earlier than the *modus tollens* ($[(a \supset b) \,.\, \bar{b}] \supset \bar{a}$); the combination of hypothetical with disjunctive later still; and inductive arguments last of all. But once again we ourselves believe that these differences arise simply from the fact that the latter arguments are more complex and involve notions that are more abstract.

but in the code words there are three long ones. The brightest children therefore usually begin with the shorter words and pick out the first in the list—*uppa*. They argue: "This must be either HOOD or PEER. If it is HOOD, then $p=$O, and *pram* must begin with O and *help* end with O; but they do not; hence *uppa* is PEER. In that case *pram* must be E( )R( ); and this is confirmed by EARL". A similar schema is again applied when he comes to *crwgpa* and *trygpa*, which must be either BAXTER or MASTER: there the crucial letter is *y*, which may be either S or X. And this is decided by completing the missing name as NELSON.

The predominantly analytic type of child, especially the unobservant but assiduous workers among the older pupils, generally fall back on what may be called a universal working method—one that is applicable to nearly all code-problems: this consists in neglecting hunches and the like, and mechanically picking out in order (the order in the list or of the alphabet) all the letters appearing in (*a*) the code and (*b*) the clues, and noting the number of times each appears and its place (often according to its position in the word) and then starting with the most frequent: some even write out the letters in lists of their own. They thus discover that in the code there are 7 *p*'s while in the clues no letter occurs 7 times. This brings many to a standstill. But the brighter children rightly assume that the missing letter must be contained in the name of the unknown ship. Most of those who identify *p* in this way then fill in the names for i, vi, and viii straightaway. Next they usually perceive that *a* and R both occur 4 times (this solves iii), and that *r* and A occur 3 times; and, after this, the rest of the names are usually filled in by guessing or trial and error.

The dullest of the analytic types generally worked with whole words, taking each code word in order as it occurs in the list, and matching it in turn with the clue-words. This requires a good deal of patience; and many give up after two or three failures.

On the other hand, those who had had some little experience with code tests, crossword puzzles, and the like, would often try schematic methods of solution which they had encountered in the past: e.g. one child tried to rewrite all the clues, putting down the next letter in the alphabet after (or before) the one given. Two tried the assumption that consonants always represented vowels and vowels always represented consonants. Though not very appropriate, both methods, if continued far enough, will lead to a partial success. Here again the choice is probably not quite blind; but what precisely are the mental processes that lead the better reasoners usually to select almost at once a practicable scheme it is difficult to say. Experience and habit count for a good deal; but the difficulty is to discover precisely what habits or what experiences are most helpful.

Of children belonging definitely to the intuitive or non-analytic type, those who succeeded most rapidly seemed to be guided by what James would call 'sagacity' and other writers 'inspired guessing'. In some cases some logical plan of operation—e.g. the hypothetico-disjunctive—appears to be applied almost unconsciously: only since the child apparently omits those alternatives that would be eliminated by their negative corollaries, the actual schema is not easy to trace, and good luck may play an appreciable part. Here, however, there is obviously a need for more intensive research.

Practical Corollaries. Most of those who are likely to undertake research on the problems here raised will doubtless be, like their predecessors, either teachers or psychologists with an interest in educational applications. May I therefore end by indicating some of the practical deductions from the conclusions drawn?

1. It now seems plain beyond all question that under suitable conditions children are capable of logical reasoning at a much earlier age than teachers have hitherto assumed, and there are many indications that the capacity itself

may be trained. Burt, indeed, has described in some detail " how habits of reasoning can in some measure be taught by familiarizing the child with the commoner logical schemes and practising him in their use " [22][1]. Accordingly, teachers should make much fuller use of these latent capacities and endeavour to provide the conditions suitable for their development and training.

2. At different ages certain types of inference appear much easier than others, and certain fallacies much commoner. Several instances have been given above. But detailed knowledge is still very incomplete ; and the results already cited need to be confirmed and extended by further investigation.

3. Many of the schemata that the child spontaneously applies seem to be derived from, or at least greatly influenced by, what he has been taught, explicitly or incidentally, in regard to grammar[2]. Current grammatical analysis, however, is based on early scholastic logic, which reduced all reasoning to the subject–predicate form. Its most conspicuous defects arise from its failure to recognize the category of ' propositional functions '. Some of the fallacies and failures in the children we have tested are definitely traceable to these influences. Hence there is a pressing need for the whole theory of grammar and syntax to be modified in the light of recent studies of logic as language.

4. We have seen that, in children of high ability, the development of reasoning not only appears earlier, but also follows a somewhat different course. This is a further ground for providing a separate type of curriculum and separate methods of education for the brightest members of the school population, and including something like a training in logical thought. Our own studies have been concerned mainly with the processes of the average or normal child ; similar researches need urgently to be undertaken with the supernormal.

[1] In the paper that follows we hope to give the results obtained by actual experiments along these lines.

[2] Educational textbooks frequently quote Lindley Murray's dictum that " grammar is the *logic* of the elementary school " (e.g. Garlick's *New Manual of Method*, p. 277, his italics). If so, it is an obsolescent logic. In analysing such a sentence as " Tom plays football ", nearly every teacher will say that ' Tom ' is the ' subject ' and ' football ' the ' object '. Traditional logic insists that the predication must always be expanded into the form : " Tom is a footballer " (see, for instance Joseph, *Introduction to Logic*, p. 162). But this will not do if the statement was intended to answer " What is Tom doing now ? " And if the question was, " Is there anyone here who plays football ? " ' football ' would be described as the ' subject ' by those not already sophisticated by ordinary grammar. From the standpoint of logical analysis, however, both ' Tom ' and ' football ' are really the ' subjects ' of a *functional relation*. With the traditional formulation a child who wished to justify the obvious inference from such sentences as " Tom is taller than Jim; Jim is taller than George " would be forced to twist them into propositions asserting class-membership, " Tom is a person-who-is-taller-than-Jim ", and " Jim is a person-who-is-taller-than-George ": whereas of course the child's actual reasoning turns, not on a relation of class-membership, but on the 'comparative' relation ' taller-than '. Moreover, the traditional grammar leaves the pupil very confused about the nature of a ' converse ' function or relation. The converse of ' taller ' is ' shorter '. But from the statement ' Tom is a person-who-is-taller-than-Jim ', it would be quite impossible to draw the ' immediate ' inference ' Therefore Jim is shorter than Tom '. And it is precisely over the various forms of ' immediate ' inference that children get most easily confused. These are merely a few illustrative examples of the shortcomings of traditional grammar.

APPENDIX

The following are examples of the individual tests which we found most helpful in our work.

1. *Analogies.* For the Analogies test, both in its simple four-item form and in the more elaborate version which Burt calls a ' Matrix test ', the items were not (as in the verbal group-tests) mere words, i.e., just *names* of objects to be compared or selected, but the actual objects themselves—e.g. shapes, colours, or simple designs, drawn on paper (as in Burt's examples), or cut out in cardboard or blocks of wood. With a supplementary set of alternative items to choose from, the child is then required to complete the pattern or scheme set out before him, by discovering the specific structural principle on which it is based (cf. [30]): for instance, he may be required to construct (i) a ' rectangular matrix ', e.g. a 5×5 square of smaller squares, with the rows graded from blue through green to yellow and the columns from a pale shade to a dark shade through intermediate shades; or (ii) a ' triangular matrix ', i.e. a triangle of coloured squares arranged in a symmetrical pattern (e.g. first row—red; second—blue, yellow, blue; third—green, orange, orange, green; and so on); or (iii) a ' rhombus (or diamond) matrix ', consisting of two triangles placed base to base. The clues to the pattern required consist of a few pieces already inserted in their correct places. In the easier tests the shape is indicated by the shape of the tray into which the insets are to be placed (rather like the square and triangular trays used for the old fashioned ' form-boards '); with the harder tests he has to discover the shape as well as the other structural principles involved; and the pieces are merely placed on a sheet of suitably squared paper. With any given pattern the difficulty can be varied by varying the number of pieces already supplied, and even arranging some of them in a row or column. In this, as in all the other tests, only the simplest possible words are used in explaining the child's task.

2. *Mirror Test.* In a room, not too brightly lit, a strong light, shaded so that its beams shine mainly in the direction of the examiner and examinee, is placed close to the wall at the height of the child to be tested. The child at first stands against the wall about 10 ft from the lamp. The experimenter explains that he is going to send him a message in code; and with a small mirror (e.g. a shaving mirror) he throws a few short flashes of light into the child's eyes, and asks him to count them. He then demonstrates that his knowledge enables him to flash a light exactly in the direction needed and at the moment required, and that he can do so no matter where he stands: (he changes his position so that the angle of incidence is varied). The fact that there are definite rules or ' laws of reflection ' is *not* mentioned. The child is then asked to try to make signals in the same way, and (since his first efforts will generally show a good deal of fumbling) to discover the correct way of doing it. For this purpose a small piece of paper is pinned on the wall to represent the boy's face, and the boy takes up the experimenter's position. The experimenter counts the flashes aloud, and records the procedures adopted by the boy, asking him to explain the method used and the conclusion reached. He then gives him a mark for his final success (see p. 151).

3. *Ball and Wall Test.* The problem and procedure are much the same as for the Mirror. The boy kneels facing the wall at a distance of about 6 ft; the experimenter kneels on his right at a distance of about 10 ft from the boy, varying his position slightly as the experiments proceed. The boy is told to roll the ball against the wall so that it will bounce back into the experimenter's hands held ready on the floor. (After the boy had done his best with the Ball and Wall test, we usually reverted to the Mirror test to see how far what he had learnt during the former was ' transferred ' and helped to improve his method and his explanations in the Mirror Test: however, the results of this phase of our experiments will be discussed in a later paper.)

4. *Code Test.* The boy is told that a document containing the following list has been found in the possession of an enemy spy: " 1. *help*, 2. *mind*, 3. *pram*, 4. *trygpa*, 5. *dpmynd*,

6. *uppa*, 7. *onnh*, 8. *crwgpa*'', and that it is suspected that these eight unintelligible words are the names of certain ships of war in code. '' The numbers in front of each '' (the experimenter says) '' are known to indicate certain harbours round the coast where the ships are stationed; and the officer examining the spy believes that seven of the ships in question are the following: BAXTER, DUKE, EARL, HOOD, MASTER, LION, PEER. The name of the eighth is uncertain.'' The boy is then asked to try and solve the code and to explain his procedure as he goes along (see p. 154).

References

[1] ARMSTRONG, H. E. (1903). *The Teaching of Scientific Method.* Macmillan.

[2] MEUMANN, E. (1907). *Experimentelle Pädagogik.* Leipzig: Engelmann.

[3] RUSSELL, B., and WHITEHEAD, A. N. (1910). *Principia Mathematica.* Cambridge University Press.

[4] BURT, C. (1911). Experimental tests of higher mental processes and their relation to general intelligence. *J. Exp. Ped.*, I, 93–112.

[5] BURT, C. (1919). The development of reasoning in school children. *J. Exp. Ped.*, V, 68–77, 121–127.

[6] BURT, C. (1921). *Mental and Scholastic Tests.* P. S. King.

[7] SELZ, O. (1924). *Die Gesetze der Produktiven und Reproduktiven Geistestätigkeit.* Leipzig: Englemann.

[8] STEVANOVIC, B. (1926). Mental processes involved in judgment. *Brit. J. Mon. Sup.*, IV.

[9] FOWLER, H. L. (1928). The development of concepts. *Brit. J. Educ. Psychol.*, I, 13–39.

[10] MENON, V. K. K. (1929). An Experimental Study of the Mental Processes involved in Reasoning. M.A. Thesis. University of London Library.

[11] BOARD OF EDUCATION (1931). *Report of the Consultative Committee on the Primary School.* H. M. Stationery Office.

[12] STAYNOR, E. V. (1932). The Psychology of Reasoning with Special Reference to Educational Problems. M.A. Thesis. University of London Library.

[13] MURDOCK, J. H. (1933). Analysis of Reasoning Ability in School Children with a View to Determining the Group Factors involved. Ph.D. Thesis. University of London Library.

[14] NAHAPIET, K. (1933). Reasoning in Children from 7 to 11 Years of Age. M.A. Thesis. University of Birmingham Library.

[15] CLAPHAM, J. H. (1935). The Ability to Arrange Concepts and the Relation of this to Reasoning. M.A. Thesis. University of London Library.

[16] BOARD OF EDUCATION (1935). *Report of the Consultative Committee on Infants Schools.* H.M. Stationery Office.

[17] WOODGER, J. H. (1937). *The Axiomatic Method in Biology.* Cambridge University Press.

[18] THURSTONE, L. L. (1938). The perceptual factor. *Psychometrika*, III, 1–17.

[19] THURSTONE, L. L. (1938). *Primary Mental Abilities.* Chicago: University of Chicago Press.

[20] HOLZINGER, K. J. and HARMAN, H. H. (1938). Comparison of two factorial analyses. *Psychometrika*, III, 45–60.

[21] HOLZINGER, K. J. and SWINEFORD, F. (1939). A study in factor analysis: the stability of a bifactor solution. *Supplementary Educational Monographs*, No. 48. Chicago: Univ. Chicago Press.

[22] BURT, C. (1939). Formal training. *School Science Review*, No. 81, 653–666.

[23] WOLFLE, D. (1940). *Factor Analysis to* 1940. Chicago: Univ. Chicago Press,

[24] Gupta, S. B. (1942). *The Reasoning of Children aged* 7. M.Ed. Thesis. University of Leeds Library.
[25] Reichenbach, H. (1946). *Elements of Symbolic Logic.* New York: Macmillan.
[26] Shukla, M. M. (1948). A Study of Inductive Reasoning in Children. M.A. Thesis. University of London Library.
[27] Nakar, S. E. (1949). A Factorial Analysis of Reasoning Ability. M.A. Thesis. University of London Library.
[28] Burt, C. (1949). The structure of the mind. *Brit. J. Educ. Psychol.*, XIX, 161–175.
[29] Burt, C. (1950). Group factor analysis. *Brit. J. Statist. Psychol.*, III, 40–75.
[30] Burt, C. (1950). Theory of mental testing. *Yearbook of Education.* Evans.
[31] Guilford, J. P. *et al.* (1951–1956). A factor analytic study of reasoning. *Reports from the Psychological Laboratory*, Nos. 1, 3, 6, 11, 14, and 16. Los Angeles: Univ. S. California.
[32] Burt, C. (1952). Tests of significance in factor analysis. *Brit. J. Psychol. Stat. Sect.*, V, 109–133.
[33] Guilford, J. P. *ap.* (1955). *L'Analyse Factorielle et Ses Applications.* Paris: Editions du Centre National de la Recherche Scientifique.

Vol. XI
Part II

The British Journal of Statistical Psychology

November
1958

AN EXPERIMENTAL APPROACH TO THE IDENTIFICATION OF FACTORS

By Max Hamilton
Department of Psychiatry, University of Leeds and General Infirmary, Leeds

Summary

The primary object of this paper is to urge the value of factor analysis as a means of investigating psychiatric problems and to illustrate its use by an experimental study of the efficacy of a particular drug (benactyzine) in the treatment of anxiety-neuroses. At the same time it is argued that the details of the technique employed indicate a method whereby factors whose interpretation would otherwise remain obscure may be identified experimentally.

I. Introduction

It has long been recognized that the study of personality requires multivariate methods for handling the necessary data; nevertheless, except for purposes of general psychology, little use has hitherto been made of such techniques. Psychiatry has been regrettably laggard in this respect. So long as psychiatrists confined their inquiries primarily to the descriptive approach, the deficiency was not serious. But during recent years there has been a steady increase in the use of controlled trials for testing results of treatment. Yet in spite of this the statistical techniques employed have continued to be of the simplest types. When, for example, a fresh treatment, e.g. some new tranquillizer, is tried out, it is customary to rate its effects merely by means of some overall rating such as 'Recovered, Improved, Unchanged, or Worse'. This may suffice so long as the comparison is between treatment and no treatment; but it can have little value when two alternative methods of active treatment have to be compared. Different forms of treatment may give rise to qualitatively different forms of improvement; and a single rating may then conceal valuable information. Multivariate analysis of variance (analysis of dispersion) can of course be readily applied to practical problems such as these; but practical problems also have, or ought to have, theoretical significance. This paper will be concerned mainly with the theoretical implications of the experiments to be described. The techniques of factor analysis have been applied to a consideration of the qualitative results of a drug trial; and the unexpected results cast fresh light on the problems of factor analysis. They also have their clinical interest; but this has been dealt with elsewhere [4]. In the past, factor-analysts have tended to confine themselves to the isolation and identification of factors; but they have rarely *used* them. At one time it was hoped that factors would be of value in problems of selection and prediction; however, it was then found that, where factor analysis could be employed, the

method of multiple correlation could always do still better; and, where the latter was inapplicable, so was the former. In this investigation, factors will be used at different stages to interpret the meaning of the results.

II. Factor Analysis of the Data

The data were obtained from the use of a rating scale for anxiety neurosis. The scale contains thirteen traits treated as variables measured on a five-point system (see list in the margin of Table I). The mean reliability of the raters was 0·89. Differences between the means of raters were compared by *t* tests; and the mean difference proved to be 0·61 [3]. The scale was used in a controlled clinical trial of a tranquillizing drug, benactyzine. A ' double blind ' technique was used, 14 patients receiving the drug, and 12 ' controls ' receiving dummy tablets. Raw scores for the 13 variables were summed to give a total score; and the changes in such scores from the beginning of the trial to the end were compared for the two groups. The results showed that the group treated with the drug improved; but the gain was statistically significant at the 5 to 10 per cent level only.

Such a total score, obtained by summing raw scores, is roughly equivalent to the general factor of the separate variables. Two questions arise: (i) What would be the result of the trial were factor measurements employed ? (ii) If the scale is not uni-factorial, do the other factors undergo any changes under the influence of the drug ?

Measurements were available for 35 patients; and the correlations between the scores for the variables are shown in Table I. The correlations were factorized by the method of a simple summation; and five iterations were needed to secure stable communalities. Two factors were extracted, leaving only small residuals. Examination of the saturations indicated that the first was a general factor of anxiety, with positive saturations for every variable, and the second a bipolar factor contrasting psychic with somatic symptoms (cf. Table II). When the saturations are plotted on a graph, it appears that the angle between the variables is little more than a right-angle. To a close approximation therefore they can be included in two orthogonal group factors. These could be designated factors for ' psychic anxiety ' and ' somatic anxiety ' respectively. However, a specific rotation would yield indeterminate results, since it can vary by about 20°.

III. Use of Factor Measurements

Using the short method of Ledermann ([5] p.278), regression weights were calculated (Table II), and the factor measurements of the 26 subjects in the trial obtained. The changes appearing in the course of the trial were then computed, and examined by *F* tests to determine the effects of the drug. The results show that the advantage of the drug over the placebo was significant at a level just below 5 per cent for the general factor. For the bipolar factor,

TABLE I. CORRELATIONS BETWEEN TRAITS USED FOR ASSESSING ANXIETY STATES

The leading diagonal gives the estimated communalities: the upper right-hand triangle the coefficients of correlation and the lower left the final residuals.

Observed Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Tension	(·428)	·330	·537	·407	·412	·230	·328	·182	·235	−·008	·226	·054	·113	·337	·326
(2) Fears	·036	(·349)	·441	·396	·057	·294	·027	·140	−·185	−·069	−·393	−·029	−·414	·163	·045
(3) Insomnia	−·034	−·019	(·791)	·592	·483	·441	·195	·409	·078	−·063	·202	·108	·243	·383	·608
(4) Anxious Mood	−·041	−·009	·010	(·747)	·234	·561	·405	·128	−·067	−·310	−·039	−·307	·142	−·065	·136
(5) Intellectual Changes	·071	−·066	·063	−·057	(·323)	·278	·297	·081	·173	·277	·597	·009	−·133	·432	·346
(6) Depression	−·106	·126	·065	·102	−·019	(·519)	·271	·282	·055	−·030	·161	−·182	·052	−·377	·050
(7) Behaviour at Interview	·074	−·027	−·076	·088	·027	−·083	(·282)	·151	·294	−·078	·458	−·062	−·140	−·033	·156
(8) Gastro-Intestinal	−·054	·058	·112	−·041	−·158	·106	−·023	(·220)	·209	·053	·447	·202	−·009	·105	·217
(9) Genito-Urinary	·092	−·040	−·052	·015	−·091	−·031	·106	·007	(·440)	·386	·641	·411	·023	·109	·148
(10) Respiratory	−·048	·101	−·095	−·038	·118	·086	−·123	−·092	·017	(·428)	·542	·558	·135	256	·098
(11) Drug Reaction	−·029	−·153	−·019	·023	·142	−·062	·098	·108	−·035	−·026	(·900)	·349	·226	·236	·301
(12) Cardio-Vascular	·007	·079	·026	·014	−·125	·033	−·033	·067	·073	·128	−·141	(·500)	·058	·405	·263
(13) Length of History	·064	·195	−·074	·106	−·225	·145	−·084	−·098	−·006	·034	·232	−·150	(·325)	·309	·332
(14) Somatic General	·099	−·056	−·025	−·043	·261	−·245	·009	−·053	−·043	·010	·060	·026	−·081	(·566)	·479
(15) Autonomic Symptoms	−·003	−·186	·125	−·035	·078	−·030	·051	·008	−·040	−·093	·025	−·001	·038	·060	(·394)

TABLE II. FACTOR SATURATIONS AND REGRESSION WEIGHTS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Factor Saturations															
Centroid G	·356	·082	·619	·182	·311	·141	·135	·166	·181	·099	·491	·114	·091	·228	·317
Centroid BP_1	·255	·367	·321	·752	·074	·524	·220	−·004	−·344	−·587	−·466	−·619	−·133	−·311	−·098
Centroid BP_2	·070	·417	·281	·017	−·153	−·315	−·317	−·080	−·336	−·151	−·645	·054	·434	·471	·255
Group I′	·356	·035	·478	−·061	·424	·023	·185	·341	·545	·559	·842	·623	·325	·569	·523
Group II′	·543	·464	·703	·863	·371	·645	·387	·222	−·051	−·273	·000	−·329	−·56	·005	·230
Group I″	·240	−·226	·208	−·059	·429	·210	·339	·319	·637	·535	1·060	·462	−·006	·166	·260
Group II″	·401	·298	·459	·809	·292	·686	·411	·152	−·071	−·336	·000	−·463	−·151	−·265	·018
Group III″	·456	·501	·740	·302	·265	·023	·016	·217	·041	·101	·000	·268	·524	·670	·569
Regression Weights															
Centroid G	·118	−·004	·394	·097	·128	·100	·086	·075	·164	·132	—	·145	—	·115	·111
Centroid BP_1	·029	·088	·108	·433	−·022	·144	·025	−·024	−·138	−·182	—	−·242	—	−·156	−·066
Centroid BP_2	−·018	·216	·265	−·055	−·133	−·266	−·185	−·064	−·251	−·131	—	−·015	—	·293	·088
Group I′	·095	·010	·270	−·160	·109	−·028	·031	·076	·161	·201	—	·274	—	·172	·148
Group II′	·101	·075	·270	·448	·050	·141	·054	·021	−·034	−·087	—	−·124	—	−·026	·019
Group I″	·071	−·177	·029	−·094	·172	·167	·159	·098	·320	·244	—	·204	—	−·046	·042
Group II″	·068	·009	·132	·437	·062	·244	·107	·022	·005	−·082	—	−·169	—	−·199	−·054
Group III″	·070	·167	·491	·025	−·013	−·139	−·084	·003	−·080	−·008	—	·092	—	·313	·149

however, the F ratio was less than one. In this case, therefore, general factor measurements give much the same results as a total raw score (cf. Table III).

As will be noted in a later section, factorists, following Thurstone, consider that general factors of the above type are statistical artefacts, having no more meaning than averages, and that suitable rotations are necessary to obtain factors which will have a psychological meaning and be invariant in different samples and under different conditions. It was therefore decided to repeat the analysis with rotated factors.

TABLE III. ANALYSIS OF VARIANCE
IMPROVEMENT IN PATIENTS AS ASSESSED BY CENTROID FACTOR MEASUREMENTS

	Centroid factors		
	G	BP_1	BP_2
Controls ($N = 12$)			
Σx	52	44	30
Σx^2	702	392	408
Patients treated with Drug ($N = 14$)			
Σx	124	55	−46
Σx^2	1314	803	348
Results of Significance Test			
F	4·58	<1	9·80
P	<·05	—	<·005

The rotation was effected as follows. Examination of the changes produced by the drug in the assessments for individual items showed that only six had F ratios greater than 1, namely, ' Behaviour at interview ', ' Genito-urinary ', ' Gastro-intestinal ', ' Respiratory ', and ' Cognitive ' symptoms (all of which showed greater improvement in those patients receiving the drug), and ' Fears ' (which showed a greater improvement in the control group). Only the first reached a significance of less than 5 per cent probability, the others being just over 10 per cent. Since with 13 variables, this is greater than one would expect by chance, it was considered worth while examining the question further.

The original scores for these six variables were added, by making a simple sum of the raw scores of the first five and subtracting the score for ' Fears '. This furnished a new variable, which may be called ' Drug Response ', and will be included in the factor analysis[1]. It gave rise to a minor difficulty, since with successive iterations its communality increased rapidly to over 1. Accordingly, for the last iteration it was given a communality of 0·900. The

[1] Since this procedure has been envisaged from the outset, there was no need to repeat the factor analysis: the new variable, together with another, ' Length of History ', had in fact already been included.

results were then used for locating a rotation of the centroid factors by means of an orthogonal transformation matrix (Table IV).

Table IV. Rotation Matrix for Two Factors

·8328	·5536
−·5536	·8328

Regression weights and factor measurements were calculated for the rotated group factors (Table II); and the results of the drug-trial analysed as before (Table V). Both factors gave F ratios less than 1. Since this result was entirely unexpected, the investigation was carried a step further.

The residuals left after extracting two factors from the correlation matrix were further factorized, and a third factor (a second bipolar) obtained (Table II). Examination of the signs of the saturations of this factor showed that the variables ' Behaviour at interview ', ' Cognitive ', ' Gastro-intestinal ',

Table V. Analysis of Variance
IMPROVEMENT IN PATIENTS AS ASSESSED BY TWO GROUP FACTORS

	Group Factors I′	II′
Controls (N = 12)		
Σx	25	72
Σx^2	265	888
Patients treated with Drug (N = 14)		
Σx	69	117
Σx^2	581	1557
Results of Significance Test		
F	<1	<1

' Genito-urinary ', and ' Respiratory ' symptoms all lie on one side of the factor. These are the variables that showed improvement in the patients treated with the drug. The one item that showed less improvement in the patients treated with the drug, viz. ' Fears ', lies on the other side of the factor. It seemed therefore not unreasonable to label this factor ' Drug Response ' (D.R.); there is no other obvious meaning to be attached to the arrangement of the factor saturations. The name proposed was amply justified by the factor measurements obtained for this factor in testing the results of the drug trial. The figure measuring the benefit received by the group treated with the drug as compared with the control group proved to be fully significant when tested by the F ratio (level of significance well below 0·005: cf. Table III).

Once again, it was decided to test the effect of rotating the centroid factors to obtain group factors, using the vector of D.R. to pin down the rotation. Here, however, a difficulty arises. One factor can be rotated to lie along D.R.,

but the other two can still rotate freely in the plane normal to it. The problem was tackled by rotating around the first bipolar factor as an axis, until the general factor lay on the projection of D.R. on the plane defined by the general factor and the second bipolar. Using $B.P._2$ as an axis, the next rotation was made at right angles to the first until the general factor lay along $D.R^1$. The complete rotation matrix is given in Table VI. The new factor saturations are given in Table II.

Table VI. Rotation Matrix for Three Factors

·6610	·3234	·6771
−·4394	·8983	·0000
−·6082	−·2975	·7359

Factor weights and factor measurements were then calculated for the three group factors (see Table II); and the results of the drug-trial tested with them. The F test demonstrates that the first group factor (which lies along D.R.) shows the advantage of the drug at a significance level of less than 0·005. The other two factors gave non-significant F ratios. (Table VII).

Table VII. Analysis of Variance
improvement in patients as assessed by three group factors

	Group Factors		
	I″	II″	III″
Controls (N = 12)			
Σx	−5	48	64
Σx^2	489	634	458
Patients treated with Drug (N = 14)			
Σx	86	99	51
Σx^2	770	1351	297
Results of Significance Test			
F	9·12	1·45	1·94
P	< ·01	> ·20	·20—·10

IV. Discussion

Burt has consistently upheld the view that factor analysis is primarily a method of classifying variables according to the empirical relations obtaining between them [1]. Thurstone assumes that it may be used to find fundamental constants or operators, underlying, and manifesting themselves through, the observed variables [6]. The two views are not incompatible; but, since classifications are to some extent arbitrary, Burt's view emphasizes the fact that the factors obtained are also apt to be somewhat arbitrary. As he points

[1] The reader will find the description easier to follow if he illustrates it for himself with the aid of three pencils.

out, their arbitrary nature arises mainly from the arbitrary decisions made in choosing variables to be included in the analysis at the very outset of the research. Hence, in his opinion, the real purpose of carrying out a factor analysis is to secure an objective verification for the classificatory hypotheses implicitly or explicitly adopted in planning the inquiries.

Thurstone's objections were put in the form of a criticism of centroid factors. He alleged that they were statistical artefacts having no more ' reality ' than an average. Now the ' reality ' of an average depends on the process of averaging used; and that in turn chiefly depends on what we select for averaging. A price index based on chalk and cheese would have but little meaning or value. In the same way, a factor analysis of the correlations between a random collection of heterogeneous variables—" shoes and ships and sealing wax, cabbages and kings "—could serve merely as a basis for a conversation between the Walrus and the little oysters [2].

Thurstone proposed to eliminate the artificiality by rotating to simple structure. For this he was prepared to use non-orthogonal rotations, giving correlated factors. It has been said that what was thus thrown out of the door returned through the window: for correlated factors in turn give rise to a second-order factor, and this is virtually the general factor of the centroid or summational analysis under another name. That, however, is not the only objection to Thurstone's procedure. Arbitrariness cannot be eliminated merely by rotating. In the first place different rotations give rise to different factors. which are equally valid even when Thurstone's criterion of simple structure is adopted; and secondly, rotations cannot supply meanings that are not already there. The meaning of the Walrus's conversation cannot be found by any examination of his subject-matter; what is required is the Eldest Oyster's theory.

In the present experiment, factor analysis by ' simple summation ' extracts orthogonal common factors. The first or ' general ' factor shows that the variables can be regarded as forming an homogeneous group or family of variables measuring anxiety neurosis. Moreover, the family can be divided into two groups, one of psychic symptoms and the other of somatic symptoms. And these two groups can themselves be further sub-divided; so far, however, the nature of the sub-division is not altogether clear. But at this point the experiment in which benactyzine is used comes to the rescue, and demonstrates the nature of this last classification. Thus two completely different types of investigation have combined to yield a satisfactory result.

Unfortunately, a few minor difficulties still remain. From Thurstone's point of view, no doubt, the result would be almost as satisfying, and might at first sight seem to confirm his arguments. But this holds only so long as we consider an analysis into three centroid factors. With two or with three factors we can rotate according to his criterion, and the results will seem quite satisfactory. The factor of D.R. can be found either in a space of two dimensions or in a space of three. But, when we test the results with the external criterion,

only one is found to be valid. This is so, quite apart from difficulties about the precise nature of the rotation, which with Thurstone would usually be oblique. In the present case, however, an orthogonal rotation has given a fair approach to simple structure.

To return to the centroid factors: it is true that the second bipolar factor has been given a meaning by an external criterion; but there is no apparent reason why it should have that meaning. In another field, or with a different drug, the ' response to drug ' vector might have only a low correlation with the centroid factor, so that a rotation would then be indispensable. The difficulty here arises from the absence of any explanatory theory. However, we do not need a factor analysis to tell us that we are still quite ignorant as to how benactyzine affects mental functions. What the factor analysis has done is to dig out information from what might at first glance have appeared a null experiment. If the findings are confirmed on another group of patients (and this is necessary, for the groups were small and ratings on individual variables are subject to great error), then we should have an important clue for the direction of further research.

The findings reached draw attention to yet another problem: how many factors should be extracted from the given matrix ? Consider the case of a set of highly reliable and valid tests given to a sufficiently large number of subjects. Then a factor analysis by the method of principle components will give as many factors as tests; and there is no reason why any of them should be dismissed as not significant. Any other method of factor analysis would yield fewer factors. Tests of statistical significance, it would seem, are not likely to be helpful here. In the present case, with so few subjects, there can be little doubt that the third factor is not statistically significant. Yet the factor measurements are empirically useful, and reveal highly significant changes resulting from the experiment with the drug.

From an experiment on this restricted scale it would be unwise to draw any final conclusions. Nevertheless, at the very least it may be fairly said that the results of the investigation are consistent with the following general principles: (i) factor analysis has the great advantage of ordering data regarding the relations between the variables quite objectively, and in such a way that they can be readily understood; (ii) factor analysis can also be useful in replacing a set of correlated variables by a set of uncorrelated variables; (iii) it cannot be used for finding underlying variables or operators, since these must be confirmed by independent methods, which themselves do not require factor analysis to support them.

V. Summary

When a multivariate rating scale is used for assessing the results of treatment, it is customary to use the sum of the raw scores as a measure of the patient's condition. This is approximately equivalent to a general factor score. In this enquiry, factor scores have been computed both for centroid factors and

group factors obtained by suitable rotations. The factor scores so obtained have been used for testing the results of treatment in a drug trial.

Using an external criterion, it is possible to rotate the factors to give one which corresponds with the effect of the drug. This factor, when located in a space of two dimensions does not give measurements which show the effect of the drug, but in three dimensions it does. The implications of this finding are discussed from the point of view of factor analysis [1].

Acknowledgments

I have to thank Professor G. R. Hargreaves for permission to publish, and for taking part (together with Dr. J. M. Roberts) in the rating of the patients. The Research Fellowship is in part provided by the Mental Health Research Fund, to whom my thanks are also due.

References

[1] Burt, C. (1940). *The Factors of the Mind.* University of London Press.
[2] Carroll, L. (1872). *Through the Looking Glass.* MacMillan.
[3] Hamilton, M. (1958). The rating of anxiety states. *Brit. J. Med. Psychol.*, in the press.
[4] Hamilton, M. (1958). Components of anxiety and their response to benactyzine. *J. Ment. Sci.*, in the press.
[5] Holzinger, K. J., and Harman, H. H. (1941). *Factor Analysis.* Chicago: University of Chicago Press.
[6] Thurstone, L. L. (1947). *Multiple Factor Analysis.* Chicago: University of Chicago Press.

Vol. XI
Part II

The British Journal of Statistical Psychology

November
1958

THE INHERITANCE OF INTELLIGENCE AND ITS SOCIAL IMPLICATIONS

By J. Conway
University College, London

It is argued that the wide differences between the intellectual achievements of children from different socio-economic classes are due not only to environmental but also to genetic differences. In spite of the marked improvement in social conditions during recent years, the differences between the average levels for each class remain much the same as before the war. The distribution of genetic differences displays two significant characteristics : (*a*) the average level of children from the lower occupational groups is lower than that of the others, but (*b*) owing to their numbers and the wide range of individual variation, they provide a larger number of able children than the higher groups. A multifactorial process of inheritance combined with the effects of social mobility would adequately account for all the important features observed.

Recent critics have questioned the validity of this explanation, and in particular have doubted the evidence for mental inheritance (notably that which rests on the study of identical twins). An attempt is made to answer their objections in the light of newer evidence now available.

I. Genetic Factors in the Differentiation of Social Classes

Class Differences in Intelligence. The problem of the inheritance of general intelligence still excites considerable controversy, particularly in connection with its alleged effect on the allocation of children from different social classes to the grammar school at the age of 11 plus. The purpose of this note is to answer some of the criticisms more recently brought forward.

In one of the earliest researches on the measurement of intelligence, Burt [1] found unexpectedly wide differences in average performance between pupils at an ordinary elementary school and boys from the higher professional classes (sons of bishops, Fellows of the Royal Society, University professors, and the like) attending a preparatory school. Now it is often argued that, since intelligence tests were invented by psychologists who are members of a privileged academic class, the results will, almost inevitably, favour children brought up in homes of a similar type. But in this case nearly all the tests were non-verbal; nor were the boys from the elementary school in any way handicapped by poverty or ill health. Moreover, later investigations showed that children from poorer homes, who had actually been diagnosed as mentally deficient on the ground of excessive backwardness in reading, writing and arithmetic, often proved to have perfectly normal capacity when examined with tests of intelligence, while others who had failed in the ordinary scholarship examinations had abilities fully equal to those who had been successful. Finally, in the course of systematic

surveys of children attending schools under the London County Council, it was found that, when classified according to parental occupations, the average I.Q. varied widely, ranging from 89 (for children of unskilled workers) to 120 (for children of the higher professional classes) [19, 20]; and in a similar investigation, carried out with adults and designed to obtain a basis for vocational guidance, still larger differences were found for men and women in actual employment [6][1].

Differences within Social Classes. At the time of these earlier enquiries the idea of inherited differences in mental ability was generally accepted. What was not so clearly realized was that, according to the new Mendelian theory, wide individual variations about the general average would be the inevitable consequence of the recombination of the genes transmitted by the parental stocks. This was confirmed by a more detailed analysis of the test results. And from time to time children from the poorest occupational classes were discovered who had abilities equal to, and even exceeding, the average level of those in the highest social classes. By arrangement with the London County Council 'special scholarships' were awarded to many of these pupils; and their after-histories showed plainly that in nearly every case the award was fully justified.

During the inter-war period, with the aid of teachers and University officials, a fresh series of enquiries revealed that, with the methods of selection then in vogue, a large proportion of able children coming from the poorer social classes had been entirely passed over. It became clear that, although the average level of intelligence in the lower groups was well below that of the general population, nevertheless the individual variations were so wide, and the groups themselves so large, that the actual number of bright children was considerably greater than that of bright children in the professional and other higher classes; and an analysis of the figures available demonstrated that, during the period in question, scarcely half the ex-elementary pupils who were capable of a university education actually succeeded in obtaining it [19].

The Origin of Class Differences. All this introduced fresh complications into the problem of explaining the apparent differences between different social classes and accounting for their origin. Evidently if a child whose family belonged to the higher economic grades did not possess the ability needed for an administrative or professional appointment like that of his father, he would not of necessity drop to the bottom of the social ladder: wealth was inherited as well as ability. On the other hand, if a child whose parents were in ill-paid jobs did possess the requisite ability, it by no means followed that he would inevitably rise into a profession corresponding with his true capacity: poverty might compel him to seek work at the age of fourteen, instead of proceeding to a secondary school and a university. It therefore became essential to consider, not only variations in innate ability, but also the actual extent to which children born in one social class might rise or fall to another social class. The pioneer studies of Bowley [2], of Marquis and Chapman in Manchester, Saunders in

[1] A history of these and other inquiries on the subject will be found in [19].

Liverpool, and Ginsberg in London [7, 11]—an elaborate enquiry in which members of University College, the London Day Training College, and other institutions took part—furnished provisional figures for the amount of social migration; and Burt's schematic theory of the origin of class differences sought to take into account both of the two main factors that appear to be involved.

The criticisms to which the theory has of late been subjected seem due partly to a misunderstanding of the type of mental inheritance that it postulates and partly to the marked changes in social conditions that have taken place during recent years. Perhaps therefore it will be helpful to start by summarizing the argument on which it rests, slightly simplifying the original figures and the basic assumptions, in order to make the reasoning a little easier to follow[1].

The Genetic Classification. As in other fields of modern genetics, the evidence for the theory is mathematical: the hypothesis suggested is stated in the form of tentative postulates; their consequences are calculated; and the crucial question therefore is—do the results so deduced agree with the observable facts, or is there some alternative set of postulates which will yield a closer fit?

The basis of the calculation is the principle, which may readily be deduced from the multifactorial theory, that "if a number of genes combine, in the manner described above, to determine a set of graded measurements, and if these measurements are suitably grouped to yield classes instead of continuous variates, then the frequencies to be expected will be similar to those which can be deduced from a single pair of genes, for which the hybrid state (*Aa* or *aA*) is intermediate" ([25], p. 172, cf. [26], pp. 109f.). The principle allows us to classify all individuals into three main types having three different levels of average ability. For purposes of computation therefore we take a hypothetical population that is normally distributed with a standard deviation of 15 I.Q., and treat it as composed of three broad classes containing 25, 50, and 25 per cent respectively of the entire number. The lines of subdivision will then fall at the median and the quartiles, i.e. approximately at 90, 100, and 110 I.Q., and the means for the tails will be given by the familiar formula

$$m_t = \bar{x} \pm (z/q)\sigma_x = 100 \pm 1{\cdot}271 \times 15 = 81 \text{ and } 119,$$

say 80 and 120 I.Q. approximately (cf. [25], p. 172, footnote 1).

The Social Classification. Now, let us suppose that, at the beginning of its recorded history, the community in question is separated into distinct social groups, but that at first these groups bear no relation whatever to the classification in terms of mental ability. We can, for example, imagine that at this initial stage class I is a ruler class which owes its position, not to superior intellectual capacity, but merely to physical force or the hazards of war, and that the remainder of the

[1] Burt's formulae are derived from Fisher's early memoir on Mendelian inheritance as applied to man [3]. His mathematical arguments keep far more closely than I have done to the procedure outlined by Fisher in his work on natural selection [8], and in addition make due allowance for differences between the two parents and for differences in fertility and other complicating factors which are not essential to the central argument. The main effect of such factors, however, is merely to alter the rate of change and render it a little less uniform.

population forms a subject class in which one important subgroup (class III say) consists of the semi-skilled and unskilled manual labourers. As the community settles down to peaceful conditions, the work of government, administration, education, and the like, will at the outset be carried out by the members of class I. But, as time goes on, the barriers will be lowered; and there will be an increasing infiltration of abler individuals from the subject classes. At the same time (or more probably after an appreciable delay) the less competent stocks from class I will begin to sink slowly to the ranks of the humbler individuals.

In the course of centuries, under a fairly liberal government, we might expect a threefold division not unlike that assumed by Ginsberg and other members of the London School of Economics in their studies of social mobility at the present time. They refer to their three main groups as 'professional', 'salaried', and 'manual' respectively—terms which may perhaps be misinterpreted if taken as anything more than mere labels [7, 11, 24]. Ginsberg's class I, as he points out, roughly coincides with the highest class in the Registrar-General's five-fold classification, and his class III to the lowest three lumped together. A slightly better classification is that developed by Hall, Moser, and Caradoc Jones ([24], pp. 29f.), which distinguishes seven grades. The first four include all non-manual occupations with the exception of sheer routine work; together they correspond with Ginsberg's class I (though his is perhaps not so wide). The last two (a little narrower than Ginsberg's class III) include semi-skilled and unskilled manual work. We thus get once again a threefold division, with 29·9, 41·2, and 29·9 per cent of the male population in the several classes ([24], p. 93). This agrees very closely with the grouping adopted, and the proportions reached, in the inquiries carried out by Burt and his collaborators [6]. In what follows we shall suppose that, with our own scheme of classification, the sizes of the upper and lower class respectively are approximately the same, so that to this extent the distribution is symmetrical.

The Degree of Social Mobility. During the two generations preceding his inquiry, Ginsberg found that the proportion rising from class III to class I was about 6 per cent and 12 per cent respectively. The number dropping from class I to class III is harder to ascertain: with Ginsberg's data it averaged but little over 1 per cent, and had apparently diminished. But the low proportions seem due to the rather high level of the members representing class I (the 'professional class') in Ginsberg's original inquiry. Other investigators report larger but widely varying figures. In addition, we may remember, the ranks of the brighter members of the upper classes tend to be depleted by late marriage and relative infertility.

Burt's figures for London school children indicated that, at the time of Ginsberg's inquiry, about one in five of the brighter children whose parents belonged to his class III rose to a higher class; and a slightly smaller number of the duller children, whose parents belonged to class I, dropped to a lower class. Since our scheme is intended for illustrative purposes only, we shall adopt this proportion, viz., 20 per cent, for the movements in both directions; and, to keep

the calculations simple, we shall assume that the figure has remained constant throughout the whole of the period with which we shall be concerned. At the same time we shall also ignore any further complications that might arise from assortative mating or from differential fertility.

Now let A denote the gene for high intelligence (average I.Q. 120) and a the gene for low intelligence (average I.Q. 80), and let p denote the proportionate frequency of the gene for high intelligence (A), and q that of the gene for low intelligence (a), where $p+q=1$. At the very outset we may imagine that the proportions were exactly equal, i.e. $p=q=0{\cdot}5$, and consequently that the 'pool' of genes for high intelligence was of the same size in all classes; similarly with the pool of genes for low intelligence. Assuming random mating within each class and equal fertility throughout, the offspring in the first generation would consist of types AA, Aa, and aa respectively; and the relative frequencies of the three types would be p^2, $2pq$, and q^2, i.e. if $p=0{\cdot}5$, 25, 50, and 25 per cent respectively. To ensure that the standard deviation of the distribution shall have the usual value of 15 I.Q., we shall suppose that the AA's, Aa's, and aa's have average I.Q.'s of 120, 100, and 80, as suggested above.

Now, if in this younger generation the differences in ability tend to produce an interchange of members between different social classes, the proportions of the three types will inevitably alter. Let us suppose that they are changed from p^2, $2pq$, and q^2 to P, R, and Q, where $P+Q+R=1$ (cf. [26], p. 110). Since we are ignoring the effects of rare mutations, the changes in the genetic composition would be due solely to social instability, i.e. to migration from one class to the other.

To make the resulting calculations quite clear, let us suppose that classes I and III each contains exactly 100 younger members, and that the interchange between classes takes place along the following lines. In class III, out of the 25 individuals who are of type AA, 20 per cent, i.e. 5, would move to a higher social class. On the other hand, all the 25 members who are of type aa will remain in class III, but 5 more of this lower type will move in from one of the higher classes. We shall also suppose that analogous changes, though in the opposite direction, take place in class I. Keeping for the moment to class III, it will be seen that, as a result of the interchange assumed, genetic composition of the parents for the next generation will be appreciably altered: there will now be only 20 of type AA, 50 (as before) of type Aa, and 30 of type aa.

But the genetic composition of the children in the next generation will not (as is sometimes supposed) be identical with that of their parents. The requisite calculations must take into account two intermediate steps, and we must proceed in accordance with the method outlined by Burt ([26], pp. 109–110). First, we must calculate the frequencies of possible unions arising with random mating (*loc. cit.*, Table IV); and next we shall calculate the frequencies of the different types of child born from these unions (*loc. cit.*, Table V). We then find that the composition of class III has changed to AA's 20·25, Aa's 49·50, aa's 30·25. And generally, if with the parents the new proportions for the three types are

P, R, and Q, then the proportions for the children will be

$$(P+\tfrac{1}{2}R)^2, \quad 2(P+\tfrac{1}{2}R)\ (Q+\tfrac{1}{2}R), \quad \text{and} \quad (Q+\tfrac{1}{2}R)^2$$

respectively. These proportions would again be genetically stable; but they will be altered by further migrations from one class to another when the children grow up. Meanwhile, let us note how the effects of mobility have modified the average level of intelligence obtaining in class III: it drops from 100 I.Q. to $20{\cdot}25 \times 120 + 49{\cdot}50 \times 100 + 30{\cdot}25 \times 80 = 98$ I.Q. Thus during the first generation after the interchange begins the I.Q. will be lowered by 2 points.

In the following generation we shall assume that the proportionate amount of interchange of the various types proceeds much as before. Once again, about 20 per cent of the *AA*'s will leave class III, and a similar number of *aa*'s will enter. Omitting fractions, this means that with the new parents the genetic constitution will be changed to 16, 50, and 34 per cent; and consequently the proportions of the three types among the offspring will be approximately 17, 48, and 35 per cent. The I.Q. will therefore drop to 96·4. In the third generation it will fall to a little over 95; and so on: until by the tenth generation it will have dropped to about 90. And it will continue to decline in this way at a diminishing rate, but will never drop so low as 80.

There will be a similar progressive change, though in the opposite direction, in the average level of class I. Here, by the tenth generation, the level will have reached 110; and, though it continues to rise, it will never quite reach 120.

The Final Result. The rate of change in both the upper and the lower classes may be roughly fitted by an exponential curve. For class III, for example, a moderately good fit is in fact given by the equation $\text{I.Q.} = 80 + 20 \times e^{-0{\cdot}082t}$, where t denotes the ordinal number of the generation in question. Burt, however, assumes that the changes described would proceed relatively slowly at first, then accelerate, and eventually decline; so that, apart from incidental fluctuations, the change in the intellectual level would tend to follow a logistic (or 'double-exponential') curve, $y = (e^t - e^{t})/(e^t + e^{t})$, where y denotes the gain or loss in intellectual level and t the time that has elapsed (cf. [16], p. 645).

But apart from the possible effects of a differential birthrate (a problem which we are here ignoring altogether), the general level of the intermediate class (class II, largely skilled workers or employees in routine non-manual jobs, shopkeepers, and the like) will undergo little change. Our scheme allows for the stocks that rise from class III to class I, and those that drop from class I, to pass through this intermediate class: but we assume that the movements in and out from opposite directions tend roughly to balance one another. If they did not, then, as a result of the large size of this intermediate group, the average level of the population as a whole would itself be modified; but, in our own country at least, there is no evidence whatever that this has actually occurred to any discernible extent.

The Final Result. The final outcome with a population whose history conformed to this broad scheme seems clear. In the higher class (consisting mainly of intellectual workers) the pool of genes for high intelligence will tend to

increase, but the genes for low intelligence will never by any means be completely weeded out. In the lower class (consisting mainly of unskilled workers) the pool of genes for high intelligence will diminish, but, even apart from possible mutations, it will never be entirely exhausted.

In either case the changes will be asymptotic, and tend ultimately to an equilibrium; but of course this equilibrium might easily be affected on the genetic side by any marked alteration in matrimonial custom and on the social side by any marked environmental changes, such as improvements in the educational ladder leading from the lower class to the higher. In the population as a whole, however, neither the mean nor the variance nor the general shape of the total distribution will be altered to any appreciable degree.

II. Environmental Factors in Class Differences

Gaps in the Educational Ladder. Improvements in the educational ladder (such as those introduced by the Act of 1944) may in the future have an appreciable effect on the rate of interchange between the different social strata and so conceivably modify the genetic composition of the different occupational classes. It is therefore worth while inquiring whether any such effects are already perceptible. As we have seen, surveys carried out during the period between the two wars indicated that " over 40 per cent of the pupils from elementary schools who were capable of a university education never obtained it " ([19), p.87). The failures were commonly attributed to extraneous causes—poverty and its various concomitants, lack of cultural interests and opportunities within the home, and, above all, the methods by which children were selected first for the secondary schools (as they were then termed) and then for the universities; and it was argued that many of the failures would be obviated if (i) the general standard of living among the poorer classes was improved, (ii) education of a grammar school type was made free for all who required it, and (iii) the criterion for selection was based not merely on the results of a purely scholastic examination in Arithmetic and English, but also included a test of intelligence and other devices to detect pupils who were innately bright though lacking in mere educational attainments[1].

Most of the environmental handicaps have now been removed. The number of boys from the ' working classes ' who today enter the grammar school is in many areas more than double what it was twenty-five years ago. Nevertheless, a well marked disproportion in the figures is still discernible. On this point the most recent and most detailed information comes from the survey, published by Dr. Floud and her colleagues, for two different areas—Middlesbrough and South-west Hertfordshire [27]. In Middlesbrough 68 per cent of the candidates from the class that included ' professional workers, business managers, and owners'

[1] The notion that educational psychologists " laid exclusive stress on alleged inborn variations in so-called general intellectual ability " is not quite correct. Burt and others, for instance, laid almost equal stress on temperamental and moral factors ; the temperamental factors, it was argued, were to a large extent innate ; but the moral factors (differences in interests, ideals, personal ambition, ethical code, etc.) were, he believed, partly traditional, and so due mainly to environmental conditions.

succeeded in obtaining places in a grammar school, as contrasted with 9 per cent from the 'unskilled workers'; in South-west Hertfordshire a similar difference was found, though the divergence was not so striking—namely, 51 per cent and 11 per cent respectively.

The Average Intelligence of the Different Social Classes. It might perhaps be supposed that the vast improvement in social conditions and educational opportunities that has been achieved during the last twenty-five years would have produced a marked rise in the intellectual level of children from the lower social classes, particularly if, as is so commonly maintained, their performance at intelligence tests is itself very largely a reflection of their environmental situation. Let us therefore glance at the detailed figures and see how far the class-differences in average intelligence have altered. Table I gives the average I.Q.'s for the different

TABLE I. AVERAGE I.Q. FOR DIFFERENT OCCUPATIONAL GROUPS

Class	Occupational Category	London 1922–27	S.W. Hertfordshire 1952	Middlesbrough 1953
1	Highest administrative	120·3	—	—
2	Professional	114·6	113·0	115·2
3	Clerical	109·7	109·2	108·2
4*a*	Commercial	104·8	103·7	104·4
4*b*	Skilled	101·5	100·1	99·0
5	Unskilled	95·1	97·1	96·5
	Average for Working Class	98·3	98·2	

occupational groups[1] as reported by Dr. Floud and her co-workers; and for comparison I have inserted corresponding figures for London children about 35 years ago. South-west Hertfordshire is described as "a prosperous but socially differentiated area, providing a maximum response to postwar measures of social and educational reform": here, we are told, "the level of material prosperity *in all classes* is high enough to prevent its having any consistent influence on success rates". Indeed, as one headmaster states, the wages offered to youths from the working classes are so attractive that a large proportion of those who enter the grammar school leave before their time is up. Middlesbrough, on the other hand, is a highly industrialized area, with a large class of unskilled workers. But even among these, it is said, the majority now receive wages of £7 10*s*. 0*d*. and over.

On the whole, there is a close agreement between the class-means for the two different periods. The 'intelligence test' used by the education authorities concerned was, we understand, a version of the Moray House group tests. On applying these tests to a random sample of children, who subsequently took the group test used by Burt as the basis for his I.Q. scale in the London survey, we find that the standardization of the two is much the same: if anything, the Moray House

[1] The classification adopted by Floud *et al.* is not quite the same as that adopted by Burt and his collaborators: e.g. the latter include a 'class 5' termed 'semi-skilled', which Floud appears to have divided among the classes here labelled '4*b*' and '5'.

tests appear slightly easier, though the differences are not significant statistically[1]. In Dr. Floud's table, the unexpectedly low standard deviation for the whole group of children (14·1 in both areas instead of 15·0) suggests that an appreciable proportion of those who were 'educationally subnormal' did not actually sit for the 11+ examination. In the London survey the entire school population was tested—not only the dull and backward but also those attending schools for the mentally deficient. The omission of the educationally subnormal in the more recent surveys would also account for the somewhat higher scores obtained by children of unskilled workers. However, the figures for skilled workers show a slight decline; and this would presumably appear larger still had the educationally subnormal been included. But let us waive this minor point, and take an overall average for the group that Dr. Floud terms 'working-class': the pre-war figure is 98·3 I.Q. and the post-war 98·2 I.Q. There can be no question that the environmental conditions of working class children in Hertfordshire and Middlesex today are far superior to those of working class children in London at the time of the pre-war surveys, which, it should be remembered, include a period of marked economic depression. Yet it seems clear that the improvement in social circumstances has in no way produced a rise in the level of intelligence as assessed by the intelligence tests. Hence there seems no cogent ground for supposing that the lower averages of the children from the lower occupational groups are in any important degree due to economic handicaps.

III. Current Criticisms of the Genetic Hypothesis

Arguments for a Purely Environmental Hypothesis. So far, then, an examination of the available data seems entirely consistent with the genetic theory as at least a partial explanation of the class differences observed. Even so it would not be right to accept it as fully proved without inquiring whether there may not be other objections of a more general nature or whether some alternative explanation might not fit the facts quite as well. In a paper published earlier this year [30] Dr. Halsey (one of Dr. Floud's collaborators) has attacked the problem from this standpoint, and brought forward several arguments which in his opinion "throw considerable doubt on the possibility that mean differences between the classes in measured intelligence are innate".

If I understand his contentions rightly, they rest on two main lines of reasoning. First, it is said, even if we accept a multifactorial theory of the inheritance of intelligence, and allow for selective migration from one occupational group to another, there would still be no reason for believing that such processes, no matter how long they might be continued, could result in any significant difference in the average

[1] I may add that, so far as the limited data available go, the re-application of the older tests to pupils now attending London schools shows very little difference so far as intelligence is concerned, but a slight decline in arithmetic and spelling: on the other hand, in tests of comprehension in reading there seems, if anything, to be a slight improvement. However, before such results can be accepted, a systematic survey of carefully selected samples would be essential. Even then, it is to be remembered, there have been appreciable social changes in many areas, and nowadays families move in and out of London on an even larger scale than before.

innate ability of the different social classes at the present day. Secondly, we are told, the only convincing evidence on the problem of environment versus heredity is to be found in studies of identical twins reared apart; and here the I.Q. differences correlate so closely with differences in the social environment that, if anything, the results support an environmental rather than a genetic hypothesis.

Let us therefore examine each of these arguments in turn.

A. Social Structure and Genetic Variability

A Genetic Model. In the earlier part of his paper Dr. Halsey describes ' a genetic model of intelligence ', which at first sight closely resembles the ' multifactorial Mendelian model ' described by Burt and Howard ([26], p. 99). He accepts Burt's distinction between genotypic intelligence (γ), phenotypic intelligence (g_a, i.e. ' intelligence ' as exhibited in everyday behaviour, which of course is the complex effect of genetic, developmental, and environmental conditions), and intelligence as assessed by a particular set of tests (g_i); and renames them I_A, I_B, and I_C. He also accepts a multifactorial theory for the inheritance of I_A (or γ). He then traces the evolution of " a society having a population of 1,000 ", which is divided into two classes only: (i) a ' low class ' containing 900 individuals and (ii) a ' high class ' containing no more than 100 individuals.

For the purposes of his model he adopts " an assumption of single gene determination of I_A ". This he decribes as a ' gross over-simplification ', but he believes it " gives maximum weight to genetic factors ". In a later section of his paper he endeavours to remove what he considers to be possible fallacies involved in working with a single pair of genes by changing to two pairs (A and a; B and b); however, the main effects undergo little modification. To secure a binomial distribution with the usual standard deviation of 15 I.Q., an I.Q. value of 120 has again to be assigned to the homozygote AA, 100 to the heterozygote Aa, and 80 to the homozygote aa. The hypothetical figures proposed for the interchange between social classes, however, are somewhat unexpected. It is assumed that in each generation ten of the least intelligent members of the high class will drop down to the low class, and that ten of the most intelligent members of the low class will move up to the high class. These are proportions of 10 per cent and 1·1 per cent respectively: i.e. the downward movement is nine times as rapid as the upward. These percentages, however, seem quite at variance with what has actually happened. If, for example, we look at Table VII in Glass and Hall's study of social mobility ([24], p. 7), we shall find that the proportion of men in categories 1, 2, and 3 who had fathers in categories 5, 6, and 7 was much larger than the proportion of men in categories 5, 6, and 7 who had fathers in categories 1, 2, and 3, i.e. the upward movement was much greater than the lower. The same phenomenon was noted by Ginsberg in his earlier survey. As he observes, " there is relatively little downward movement from I to III ". It is true that during the last generation or two " there had been an increase of mobility upward ", but it is difficult to believe that the proportions have switched over so rapidly as the figures adopted by Dr. Halsey would imply.

Another improbable assumption is that the *absolute* numbers transferred have remained the same generation after generation, so that the *proportionate* number of *aa*'s who move downward to a lower class has steadily increased. The result would be (as indeed Dr. Halsey points out) that, after about seven generations, *all* the members of the *aa* type in the high class will have dropped down to the low: in other words, during this period "low intelligence (*a*) is bred out of the higher class"; and the consequence is that in the end this class consists of a homogeneous population all of whom have an I.Q. of 120. On the other hand, "high intelligence is never bred out of the low class": their mean I.Q. "decreases in a decreasing rate to 98". At the same time, "the number of homozygotes increases in each generation, i.e. there are more people of high intelligence and more people of low intelligence"; in short, the standard deviation progressively increases.

These effects of the model proposed are quite unlike what is known of the history of communities such as our own. In part the paradoxical results seem to be due to the asymmetrical subdivision that he adopts to begin with. Evidently the main reason why the higher class tends to become more homogeneous at a much greater rate than the lower class is that the latter is nine times as large. Dr. Halsey's upper class in fact consists quite literally of "the upper 10" (i.e. the upper 10 per cent). His subdivision therefore is more like that discussed by Burt in his later article, where the pre-war school population was divided into fee-payers and non fee-payers; but there the mean I.Q.'s were 112 and 98, not 120 and 98: yet there, even if we take the highest occupational category of all, i.e. those belonging to 'the higher professional or administrative' groups, in which the children *do* possess an average I.Q. of about 120, the standard deviation is still only about 1 point lower than that of the whole population (i.e. about 14 I.Q. if that of the general population is taken to be about 15 I.Q.); hence the group is far from homogeneous, and the families still include a discernible proportion of dull and even mentally deficient cases.

The natural inference from the figures Dr. Halsey thus obtains must surely be that there is something wrong with the basic assumptions. This, however, is not the conclusion he himself accepts. "The existence", he says, "of *individual* differences in a postulated I_A remains as a 'legitimate hypothesis'; but, especially if I_A is polygenically determined, their issue in *social class* differences is doubtful". And in a later note he and Dr. Floud have stated that they "take the view that the polygenetic nature of intelligence and the history of social mobility in this country virtually exclude the possibility that the present distribution of *A* [high intelligence] between social classes is such as to give the mean differences between them of the order given by studies in the distribution of measured intelligence": in short, they consider "the most plausible hypothesis to be a hypothesis of near-randomness in the social distribution of intelligence *A*".

As evidence confirming this alternative hypothesis several supplementary arguments are introduced. In the first place, it is said, "only where the variance

in a subgroup is less than 50 per cent of that of the total population can we be *certain* that both heredity and environment are at work ''. Now with the data[1] that he selects for analysis Dr. Halsey finds that the reduction in variance ranges from a figure that is 88 per cent of the variance of the total group down to 67 per cent. He therefore contends that '' the class differences in I_C [tested intelligence] *may* be *entirely* environmental '' (my italics). And from this he finally infers that ''the observed differences between social classes are more likely to be explained by environmental rather than by genetic factors ''.

The logic of this is again rather puzzling. Dr. Halsey apparently is only prepared to admit the partial influence of genetic factors if the proof shows that this conclusion is ' certain '. But in empirical science no proof can ever yield an absolutely certain inference: the most we can expect is that the conclusion proposed shall appear more probable than any other. No doubt a critic can always claim that the opposite conclusion *may* be true. But why Dr. Halsey should transform this into the statement that the opposite conclusion is ''more likely'' remains altogether obscure.

B. Evidence from Twins

Criticisms of Earlier Studies. In order to test the validity of a purely environmental hypothesis, Dr. Halsey thinks '' it would be necessary to take a random sample from newly born babies among (say) unskilled workers, distribute them at random among the other classes, and compare the mean [measured intelligence] of the ' displaced ' and ' control ' groups ''. The practicable comparisons that come nearest to such an inquiry are to be sought, so he tells us, in studies of foster children and of twins in which the cultural level of the home has been taken explicitly into account.

[1] The data used were not obtained from researches or surveys planned expressly to study the distribution of innate ability among a random sample of the school population, but were merely the results of an ' intelligence test ' so-called (a better name would be ' general classification test ') which happened to have been set by the local educational authority as part of their 11+ examination. But it is not the purpose of such a test to measure innate ability as such : its object is merely to assist in selecting pupils who, in the light of their *present* qualifications, seem likely to profit by a grammar school course. When Dr. Halsey takes 50 per cent of the variance as the proportion needed to demonstrate that ' heredity ' plays a part in determining individual variations in a measurable characteristic, he appears to have in mind Fisher's statement that the ' total contribution of parentage ' to the variance of such characteristics as stature amounts to about 50 per cent or rather more (cf. [4], p. 216). Stature of course can be measured with a negligible amount of error ; intelligence (whether innate or acquired) cannot. This raises rather a different problem. And the writers whom Dr. Halsey cites as suggesting a figure between $\frac{1}{2}$ and $\frac{3}{4}$ as the '' fraction of the variation attributable to heredity '' were in point of fact thinking of the *accuracy* with which the innate component could be assessed, not merely of its proportionate contribution as contrasted with that of environment. Finally, the problem with which Dr. Halsey is really faced seems different from either of these questions. The children as classified in his table have been sorted out and selected according to ' father's occupation ' ; but, even if ' occupation ' could be accepted as an indication of the father's *ability*, nevertheless, as Fisher points out, the contribution of the *father alone* is no more than 25 or 26 per cent (*l.c.*, p. 214) not 50 per cent (which is the contribution of ' total parentage '). Now the average reduction of variance exhibited by Dr. Halsey's data amounts to 84 per cent ; but it is ' far less in the case of occupations where the members are most severely selected for intelligence ' (the professional, and, to a less extent, the clerical groups). Moreover, much of the reduction—at least 6 per cent according to recent estimates (cf. [26], p. 126) —must be due to ' unreliability ', i.e. to sheer inaccuracy. Hence the figures Dr. Halsey reports would in fact seem to be almost exactly what we *should* expect on the genetic hypothesis outlined above.

Studies of foster children he dismisses quite briefly on the ground that the placement of such children tends to be selective. However, his references suggest that he has in mind only those investigations which have been carried out in America. In the London inquiries [19, 25] the main comparisons related to children of different parentage brought up from early infancy in L.C.C. residential institutions: here there could be no question of selective placement. But about these researches he is silent.

For the comparison of twins brought up apart he refers to a single study only (also American), namely, that of Newman [15]. Here, he points out, "the I.Q. differences correlated 0·51 with judged discrepancies in social environment"; but, he says, "we can be quite sure that even this result under-estimates the effects of social class". Why he should be so "sure" (apart from personal preconceptions) he does not explain: in any case no such inference is suggested by the details in Newman's own account. However, no matter how large the true correlation between environment and test-results, it would not follow that a superior environment is necessarily the cause of superior intelligence: a family of high innate ability might easily rise to a higher class (as he himself admits), and so create a superior environment, while handing on its high ability by direct inheritance to the children.

Much of the confusion that surrounds his arguments seems to arise from the fact that Dr. Halsey frequently speaks of differences in intelligence and differences in the scores obtained from intelligence tests as though the two were the same; and, what is worse, he apparently assumes that what is true of one type or one set of intelligence tests holds of all the rest. No doubt, when we remember the nature and the purpose of the tests actually used in the 11+ examination, we should expect them to be influenced in some degree by cultural differences. Such tests, as a rule, are largely if not wholly of a verbal type. Yet, because a bookish environment will tend to raise a child's performance in vocabulary tests or in tests of a verbal type, that does not mean that it must also increase his performance in tests that are essentially non-verbal, such as we ourselves used in our studies of twins. Nor is he quite correct when he says (of Newman's cases) that "the types of homes into which the twins were placed rarely differed very much". As Newman himself explains, with some pairs there was a difference "between relative wealth and poverty"; in others, between "cultured and uncultured family groups"; in others again "between city and country life". 'Helen', for instance, who had a Stanford–Binet I.Q. of 116[1], had been "through grade school,

[1] Anyone familiar with the researches by Hugh Gordon, Frances Gaw, and others on canal-boat and rural children would realize that the Stanford–Binet scale cannot possibly be accepted as providing a fair comparison of *innate* ability with individuals brought up in such different environments—least of all with adults. The conclusion Newman himself drew from his data (obtained with Stanford–Binet) is quite different from that drawn by Dr. Halsey. He points out that of the four cases showing a wide disparity in test-performance, the twins had been subjected to very different types of education; and he notes that "even with a good education the poorly endowed individual does not reach the level of a potentially able but poorly educated person". He concludes that, wherever there was a large discrepancy in the actual scores, it seems probable that "the twin with the lower I.Q. nevertheless had an *inherited* capacity to reach the rating of the other" ([15], p. 146).

high school, and college", whereas her sister 'Gladys' (who scored only 92) had spent much of her earlier life "in an isolated part of the Canadian Rockies, where there were no schools". And in several other cases the homes showed differences that were almost equally well marked.

Later Investigations. Newman's investigation [9] was carried out over twenty years ago; nor is it by any means the only inquiry of its kind, as the uninformed reader might have supposed from Dr. Halsey's silence. Among the American inquiries carried out in the United States, that of Barbara Burks deserves at least a passing mention: she examined five pairs of twins (four at least almost certainly monovular) who had been reared apart [18, 21]. In Holland Bouterwek has investigated a pair of twin sisters separated early in life [13]. In Copenhagen Dr. Juel-Nielsen is making a study of a dozen or more uniovular pairs brought up in different homes; and part of her work has already been published [28]. In Stockholm Dr. Husen is carrying out an analogous investigation. Other cases will be found in the early report by Müller [5]. Several researches on more general lines have also been attempted with a view to investigating the possible differences in 'personality' as a whole [17, 22]. The main results are in keeping with those already reached by Newman: "the effects of environment are greatest for personality traits, next for achievement" (some would reverse these two), "next for 'intelligence', and least of all for physical characteristics". And the final conclusion to be drawn from all these different researches fully confirms that stated by Harold Carter in his review of 'Studies of Twins Reared Apart' in the American *Yearbook* [14]: "the whole array of cases suggests that environmental differences between homes are less effective for the modification of I.Q.'s than proponents of nature theories would like us to believe: drastic differences in educational and social environment are at times associated with only moderate differences in I.Q.'s; and every increase in the body of data brings more evidence of hereditary influence"[1].

Since Dr. Halsey is dealing with the problem of class-differences in Britain, one would have expected him to refer to the researches on twins carried out in this country, especially as some of them relate to much larger groups than the American. Others, it is true, refer only to one or two cases; yet their cumulative effect is appreciable.

One of the earliest and most instructive studies was that of Saudek on 'a British pair of identical twins reared apart' [12]; but this was confined to a single pair. More recently, Dr. H. J. Eysenck, a former member of the Department of Psychology, University College, has supervised two instructive researches carried out at the Maudsley Hospital. In the first, a number of twin school children were

[1] It should be noted that we are not here concerned with results derived from the study of twins generally. On this a vast literature is available: see, for example, Hermann, L. and Hogben, L. (1933), 'The intellectual resemblance of twins', *Proc. Roy. Soc. Edin.*, LIII, 105–129, and Hogben, L. (1931), *Genetic Principles in Medicine and Social Science*, ch. I, 'The Problem of Twin Resemblance' and refs. Hogben's conclusion is that "a survey of the existing data interpreted in the light of modern biological theory justifies Galton's belief that the study of twin resemblance provides clear evidence of inborn differences in capacity to respond to a given social environment" (*l.c.*, p. 34): see also Hirsch [10] for further references.

investigated at the Genetics Unit; and, with the set of intelligence tests regularly employed by the Unit, a correlation amounting to 0·76 was found between pairs that were monozygotic [23]. This result served as a kind of datum line for the second research carried out by Mr. J. Shields [29]. In the course of a television programme he issued an appeal for twins who had been reared apart to come forward for testing and investigation. Up to the present as many as 38 uniovular pairs have presented themselves for study—25 female and 13 male, the majority being over 20 years of age. They came from all parts of the country; 25 were separated during the first year of life, most of them not long after birth: "one was taken from his mother before his twin brother had been born". A few were parted later in life, and were retained to investigate the effects of the age of separation. About half had been brought up by different members of the same family. In twelve cases, however, the foster-parents were quite unrelated to the twins themselves. Often the environment was widely different. "One father", we are told, "sold a twin to settle his debts; the child was brought up in Chile and married a Frenchman; her sister was sent to Denmark during childhood and subsequently married an Englishman."

Nearly all were examined at the Institute of Psychiatry; and the scheme included not only intelligence tests but a study of personality. The diagnosis of ovularity was based on finger-prints, blood grouping, and P.T.C. testing. Here we are interested primarily in the results of the tests for intelligence. Two were administered: one was a non-verbal 'dominoes' test, similar to the University College 'matrices' test; the other was the 'synonym section' of the Mill Hill Vocabulary Scale. With the non-verbal test the intra-class correlation was 0·77; with the verbal, as might be expected, it was not so high (0·74). In the case of ordinary siblings brought up together the correlation for both tests, we are told, is in the neighbourhood of 0·50. The increase with the twins is not so large as that obtained in our own cases. But the lower figure for the vocabulary test might easily be explained by the cultural differences in the homes in which the different members were brought up. And a single non-verbal test, as short as the 'dominoes', is bound to have a relatively low reliability. Even so, however, the correlation obtained with identical twins reared apart is quite as high as that obtained by Mr. Blewett with identical twins reared in the same home. Accordingly, as Mr. Shields observes, it would be almost impossible to account for such figures as these unless we are prepared to assume that the capacity tested depends mainly on genetic constitution. Moreover, "if our analysis of these separated twins is correct", he says, "early age of separation is not such a vital point as critics have sometimes suggested". He believes too that his data demonstrate that "uniovular twins at times show striking similarities in important aspects of personality, even when brought up from an early age apart from each other and by mothers of different character". If then traits of personality are so plainly influenced by heredity, we may reasonably regard these further results as an *a fortiori* confirmation of a similar assumption for intelligence[1].

[1] The first report of the results was presented during the symposium on genetics in October, 1956: for fuller details see [29].

A Comparison of Correlations. Since the last review of our own cases was published [25] our collection has been still further enlarged. The cultural differences between the homes were, on the whole, even wider than those described by Mr. Shields. This doubtless is partly due to the sources from which our material was obtained. Our earlier cases were encountered during the routine inspections of children brought up in residential institutions under the L.C.C.: not infrequently it turned out that the child in the institution was a twin, and that the other twin had been left with the mother or with relatives. Among our later cases most were discovered through personal contacts; and, as a result, many of them came of educated parents, usually school teachers or members of a University staff: when the pair was separated, one twin generally remained with the mother and shared her cultural environment, while the other was boarded out, usually with persons of much lower intellectual status.

A typical example is that described by Burt in his Bingham Lecture. George[1] and his brother were children of an Oxford don who had died a few weeks before their birth. Unable with her slender purse to bring up two boys, the mother had secretly arranged for one to be boarded out. George remained with his mother; and, after a brilliant career at school, had obtained a first class degree in modern languages. His former teacher, encountering him (as he supposed) on the arrival platform at Paddington, was surprised to hear himself answered in a strong Welsh accent by a stranger who declared that his name was Llewellyn Jones. It turned out that Llewellyn was the twin who had been adopted in infancy by an elderly Welsh couple, and brought up on an isolated farm in North Wales. He was now himself a young but successful farmer. Until this accidental meeting neither twin had the smallest inkling of the other's existence. We tested both. Llewellyn's reading, and his verbal abilities generally, were those of a child of barely eleven; but his high I.Q. was practically the same as his brother's (137 as compared with 136 for George).

The number of cases of this type—all identical twins reared apart from early infancy—now amounts to 42. Only five have been brought up by relatives; and these, as it happens, include two pairs showing the largest discrepancies. For the whole group the correlation with the socio-economic ratings of the homes is admittedly positive, namely, 0·26; but the figure is only just significant. Since the data previously published (*loc. cit.*, p. 168) relate only to the smaller sample then available, I give a more complete and up-to-date set of correlations in the table with corresponding figures from Newman's inquiry (the investigation cited by Halsey) added for comparison.

As regards intelligence, it will be seen, the most significant feature is the large correlation between the assessments for identical twins even when they have been reared apart. With both Newman's data and our own, they are appreciably higher than the corresponding correlations for non-identical twins reared together. With

[1] The published names are fictitious substitutes, used merely for convenience of cross-reference.

TABLE II. CORRELATIONS BETWEEN MEASUREMENTS FOR MENTAL, SCHOLASTIC, AND PHYSICAL CHARACTERISTICS

Measurement	A—Burt and Conway						B—Newman, Freeman and Holzinger		
	Identical twins reared together	Identical twins reared apart	Non-identical twins reared together	Siblings reared together	Siblings reared apart	Unrelated children reared together	Identical twins reared together	Identical twins reared apart	Non-identical twins reared together
Mental (Intelligence)									
Intelligence :									
Group test	0·936	0·778	0·542	0·515	0·441	0·281	0·922	0·727	0·621
Individual test	0·919	0·846	0·526	0·491	0·463	0·252	0·910	0·670	0·640
Final Assessment	0·928	0·881	0·551	0·538	0·517	0·269	—	—	—
Scholastic									
General Attainments	0·894	0·629	0·831	0·814	0·526	0·535	0·955	0·507	0·883
Reading and Spelling	0·943	0·645	0·915	0·853	0·490	0·548	—	—	—
Arithmetic	0·870	0·726	0·748	0·769	0·563	0·476	—	—	—
Physical									
Height	0·956	0·942	0·472	0·503	0·536	−0·069	0·981	0·969	0·930
Weight	0·929	0·884	0·586	0·568	0·427	0·243	0·973	0·886	0·900
Head length	0·961	0·958	0·495	0·481	0·536	0·116	0·910	0·917	0·691
Head breadth	0·977	0·960	0·541	0·507	0·472	0·082	0·908	0·880	0·654
Eye colour	1·000	1·000	0·516	0·553	0·504	0·104	—	—	—

our own data, as with those of most other observers, the resemblances between non-identical twins, even when reared together, are not much greater than those for ordinary brothers and sisters. The fact that, when identical twins are reared together, the coefficients are slightly increased certainly suggests that environment may not be entirely without some effect; and, on turning to the detailed case-histories, it appears that this is usually due, not so much to similarities and differences in the cultural background, as to the amount of early ill health common to both—particularly malnutrition and infectious ailments, which were far more frequent with twins brought up by impecunious or somewhat neglectful foster parents. Somewhat similar trends are discernible in the figures for physical measurements: here the high correlations obtained by Newman for non-identical twins are rather surprising; Lauterbach's figures agree far more closely with our own (*Genetics*, X, 1925, pp. 525–68).

IV. Conclusions

To anyone who examines the foregoing data with complete impartiality it must surely be clear that genetic differences not only play an appreciable part in determining the correlations observed, but that, so far as 'intelligence' (in the technical sense) is concerned, their influence greatly outweighs that of environment. That being so, there no longer remains any reason for rejecting the view, which rests on strong theoretical arguments as well as on the empirical evidence, that the differences between the averages for different social classes are also to a large extent determined by genetic factors: in short, as Fisher puts it, "distinctions of social class are distinctions of relatively permanent biological entities, not indeed strictly insulated, but influencing one another by a constant diffusion of persons who pass from class to class" ([8], p. 211).

References

[1] Burt, C. (1909). Experimental tests of general intelligence. *Brit. J. Psychol.*, III, 94–177.
[2] Bowley, A. L. (1915). *The Measurement of Social Phenomena.* P. S. King.
[3] Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. roy. Soc. Edin.*, LII, 399–433.
[4] Fisher, R. A. (1919). The causes of human variability. *Eugen. Rev.*, X, 213–220.
[5] Müller, H. J. (1925). Mental traits and heredity. *J. Hered.*, XVI, 433–448.
[6] Burt, C. *et al.* (1926). *A Study in Vocational Guidance.* H.M. Stationery Office.
[7] Ginsberg, M. (1929). Interchange between social classes. *Econ. Journ.*, XXXIX, 554–545.
[8] Fisher, R. A. (1930). *The Genetical Theory of Natural Selection.* Oxford : Clarendon Press.
[9] Newman, H. H. (1930). Identical twins : differences between those reared apart. *Eugen. Rev.*, XXII, 29–34.
[10] Hirsch, N. D. M. (1930). *Twins : Heredity and Environment.* Cambridge, Mass. : Harvard University Press.
[11] Ginsberg, M. (1932). *Studies in Sociology.* Methuen.
[12] Saudek, R. (1934). A British pair of identical twins reared apart. *Character and Pers.*, III, 17–39.

[13] Bouterwek, H. (1936). Erhebungen an eineiigen Zwillingspaaren über Erbanlage und Umwelt. *Z. Menschl. Vererb. u. Konst. Lehre*, XX, 265–275.
[14] Whipple, G. M. ed. (1940). *29th Year book of Nat. Soc. Stud. Educ.* Bloomington, Ill., Public School Publishing Co.
[15] Newman, H. H., Freeman, F. N., and Holzinger, K. J. (1937). *Twins : a Study of Heredity and Environment.* Hutchinson.
[16] Burt, C. (1937). *The Backward Child.* University of London Press.
[17] Yoshimasu, S. (1941). Psychopathie und Kriminalität. *Psychiat. et Neurol. Jap.*, XLV, 455–531.
[18] Burks, B. S. (1942). A study of twins reared apart under different types of family relationship. *Studies in Personality Contributed in Honour of Lewis M. Terman*, pp. 35–70.
[19] Burt, C. (1942). Ability and income. *Brit. J. educ. Psychol.*, XIII, 83–98.
[20] Burt, C. (1946). Intelligence and fertility. *Papers on Eugenics No.* 2, Hamish Hamilton.
[21] Burks, B. S., and Roe, Anne (1949). Studies of identical twins reared apart. *Psychol. Monogr.*, LXIII, 1–62.
[22] Schwesinger, Gladys (1952). The effect of differential parent–child relation on identical twin resemblance in personality. *Acta. genet. med. et gemellol*, I, 40–47.
[23] Blewett, D. B. (1954). An experimental study of the inheritance of intelligence. *J. ment. Sci.*, C, 922–933.
[24] Glass, D. V. ed. (1954). *Social Mobility in Britain.* Routledge and Kegan Paul.
[25] Burt, C. (1955). The evidence for the concept of intelligence. *Brit. J. educ. Psychol.*, XXV, 158–177.
[26] Burt, C., and Howard, M. (1956). The multifactorial theory of inheritance and its application to intelligence. *Brit. J. statist. Psychol.*, IX, 95–131.
[27] Floud, J. E., Halsey, A. H., and Martin, F. M. (1956). *Social Class and Educational Opportunity.* Heinemann.
[28] Juel-Nielsen, N. and Mogensen, A. (1957). Uniovular twins brought up apart. Preliminary report of a psychiatric–psychological study. *Acta. genet. et stat. med.*, VII, 430–433.
[29] Shields, J. (1958). Twins brought up apart. *Eugen. Rev.*, L., 113–123.
[30] Halsey, A. H. (1958). Genetics, social structure, and intelligence. *Brit. J. Sociol.*, IX, 15–28.
[31] Darlington, C. D. (1958). The control of evolution in man. *Eugen. Rev.*, L, 169–178.

Postscript. The recent paper by Professor Darlington [31] provides an admirable answer, from the standpoint of the geneticist, to writers who (as he puts it) " argue, by a strange perversity, that because culture can be *socially transmitted*, it can never be *genetically created* ". Unfortunately his article appeared only as the proofs of the foregoing pages were passing through the press. Since it lends further support to the view I have endeavoured to defend, I may perhaps be permitted to summarize it here.

He considers the problem in three stages—" the Galtonian, the Mendelian, and the present or genetic stage ". After briefly referring to Burt's work on " the use of intelligence tests, especially among children ", and noting how the results " indicate that the less intelligent individuals and groups can now propagate their like faster than the more intelligent ", he goes on to consider in what way " migration modifies the intelligence of a community ", and explains in fuller detail how the more recent modifications of the Mendelian theory have " given us a new idea of human society ". On this basis he then elaborates his own " theory of the development of class structure " (illustrated by an instructive diagram). " Genetic differences " (he believes) " introduced by the mingling, friendly or forcible, of different races, sowed the seeds of social differentiation and the

cooperation of dissimilar individuals, out of which the richness of human culture has arisen." The ultimate result has been, in the more civilized parts of the world, " nations divided into differential classes ". " Inter-fertile classes, which work together without breeding together, thus uniquely distinguish advanced human societies from all other forms of animal life."

Man, however, has now started to " interfere with his own evolution by political acts and social habits ". " Governments which provide us with education forget that our parents can provide us with the ability to profit by it "; and those whose actions are " directed, not to breeding but to the present life of the individual and the present economy of the state, seem unaware of the consequences of their actions, because they are not quite certain about heredity." The final conclusion is that " the nation which takes most thought for its genetical future is most likely to have a future ".

It will be observed that the inferences here drawn are much the same as those reached by Burt from quite a different starting point in his Bingham lecture on ' The Inheritance of Mental Ability ' (*Amer. J. Psychol.*, XIII, pp. 1–14). Thus the geneticist attacking the problem from one side and the psychologist approaching it from the other have arrived at almost identical conclusions; and this of itself is surely a strong confirmation of the views they have thus both independently formed. I would only add that, since so many sociologists and educationists, like Dr. Floud and Dr. Halsey, still question the evidence adduced, the prime need at the moment is for further research and still more conclusive data. (On the issue here raised the reader may also refer to the correspondence between Dr Warburton Floud and Dr Halsey in the current issue of the *Brit. J. Educ. Pyschol.*, XXVIII, pp. 290f.).

Vol. XI Part II The British Journal of Statistical Psychology November 1958

THE RESCALING OF TEACHERS' ESTIMATES

By A. E. G. Pilliner
Department of Education, University of Edinburgh

Four methods of rescaling are described, and their advantages and disadvantages for different types of problem are discussed in detail. Each method has been applied to data obtained from a complete age-group of over 900 children. So far as they go, the results clearly show that for general purposes the method based on the ' orthogonal mean square regression line ' (i.e. on the principal axis) of the distribution, which is in fact the method in general use, is the most efficient.

I. Problem

The problem of rescaling estimates obtained for different groups (e.g. different schools, different age-groups, or the like) by means of different estimators (e.g. different tests, different judges, etc.) arises in many different contexts: for example, in the rescaling of essay marks assigned by different markers to different groups of children, of marks obtained in Teacher's Certificate Examinations by students in different Training Colleges, of teachers' estimates for junior county scholarship examinations, and, since the passing of the 1944 Education Act, for the 11 *plus* examination. This last application is of special importance at the present time owing to the increasing use now made by local education authorities of teachers' estimates.

Most recent writers on methods of rescaling for educational purposes, e.g. McClelland [7], Sutcliffe and Canham [11], McIntosh *et al.* [8], Vernon *et al.* [13], and Yates and Pidgeon [14], accept without question the usual procedure, while Sandon [9] deals with only one specific aspect of the problem. The assumptions underlying the method most commonly used, and its special merits and uses have been discussed in some detail by Greenall [4]. His terminology and algebraic proofs are not altogether easy to follow; and there would seem to be a need for a fuller and simpler discussion of the chief alternatives available, and, above all, for an empirical comparison of them by actual application to a representative composite group.

The necessity for rescaling arises from the familiar fact that the norms and the dispersions implicitly adopted by different tests or different teachers are not directly comparable. Hence a single overall test which can be administered to all the testees is needed to determine the most probable group means and group standard deviations. A regression line, based on the principle of least squares (with equalized dispersions), is commonly used for the rescaling. If the reliability and validity of the overall test were perfect, the ordinary regression of the estimates on the overall test scores could be used, though in this case the problem would be of academic interest, since there would be no need for the

estimates. Ordinarily, however, both overall test and estimates are likely to be in error: and if, as the usual method appears to assume, the error is the same for both, *a priori* argument suggests that the best line to choose will be the 'orthogonal mean square regression line' (see Burt [3], pp. 106–107, footnote 3); if, however, the assumption of equality is not justified, then the estimates and the overall test scores should in theory be weighted differently (cf. Burt [1, 2]).

It is the object of this study to examine the alternative methods which appear to be available for determining these weights, to consider the underlying assumptions on which these methods are based, and finally to compare the results obtained by applying the methods to an actual sample of nearly one thousand pupils forming practically a complete year-group. From a statistical standpoint the primary question is how to construct a single scale which shall preserve the order, and if possible the relative differences, obtained from different teachers, when each has his own standard and his own way of spreading the marks. We shall start by assuming that the adjusted estimates should approximate as closely as possible to those that would be assigned by a single hypothetical teacher or judge who was in a position to estimate the potentialities of the children in all the schools; that both teachers' estimates and test scores should furnish approximations to these hypothetical marks; and finally that within each school the teachers' estimates approximate more closely than the test scores to these hypothetical marks.

If the adjusted estimates, the hypothetical marks, and the test scores are positively correlated, and if all are on the same scale (that of the test scores), there will be some degree of correspondence between the school means and between the school standard deviations. Actually, of course, the means and standard deviations for the test scores only are available. We must therefore take these means and standard deviations as indicating approximately those of the hypothetical judge, and use them for constructing our adjusted estimates.

II. General Relations between Estimates, Test, and Criterion

We obtain first a general rescaling equation. Let z denote the teachers' original estimates, x the test scores, and y the adjusted estimates. For both x and y we may without loss of generality fix the mean for the entire sample at zero. Taking first some particular school, we let s_z and s_x denote the original S.D.'s and s_y that of the adjusted estimates. We assume (i) that within each school, $s_y = s_x$, and (ii) that the adjusted means $\bar{y}$ of the several schools are proportional to the test means; that is, we put $(y-\bar{y}) = (z-\bar{z})s_x/s_z$, and $\bar{y} = k\bar{x}$. Now, $y = \bar{y}+(y-\bar{y})$, whence the general rescaling equation is

$$y = k\bar{x}+(z-\bar{z})s_x/s_z. \tag{1}$$

The equation for the test is $x = \bar{x}+(x-\bar{x})$. Squaring and summing, first within and then over all schools we obtain

$$\sum\sum x^2 = \sum n\bar{x}^2+\sum\sum (x-\bar{x})^2, \tag{2}$$

$$\sum\sum y^2 = k^2 \sum n\bar{x}^2+\sum\sum (x-\bar{x})^2. \tag{3}$$

Moreover,

$$\sum\sum xy = k \sum n\bar{x}^2 + \sum \{r_{xz} \sum (x-\bar{x})^2\}$$
$$= k \sum n\bar{x}^2 + r_{xy} \sum\sum (x-\bar{x})^2 \quad (4)$$

where r_{xz} is the correlation between x and z for a particular school, and r_{xy} is the average correlation within schools obtained by pooling sums of squares and sums of products after the z's have been transformed into y's. If we denote the criterion by c, we have $c = \bar{c} + (c-\bar{c})$, so that

$$\sum\sum c^2 = \sum n\bar{c}^2 + \sum\sum (c-\bar{c})^2 \quad (5)$$
$$\sum\sum xc = \sum n\bar{x}\bar{c} + \sum\sum (x-\bar{x})(c-\bar{c}) \quad (6)$$

and

$$\sum\sum yc = k \sum n\bar{x}\bar{c} + \sum\sum (x-\bar{x})(c-\bar{c}) r_{yc}/r_{xc}, \quad (7)$$

where r_{yc} and r_{xc} are again average correlations within schools obtained by pooling sums of squares and sums of products. Using B, W, and T, to denote variations (or covariations) between schools, within schools, or over the total batch of schools respectively, we may abridge eqns. (2) to (7) as follows:

$$X_T = X_B + X_W \quad (2\,a)$$
$$YT = k^2 X_B + X_W \quad (3\,a)$$
$$(XY)_T = kX_B + r_{xy} X_W \quad (4\,a)$$
$$C_T = C_B + C_W \quad (5\,a)$$
$$(XC)_T = (XC)_B + (XC)_W \quad (6\,a)$$
$$(YC)_T = (YC)_B + (YC)_W = k(XC)_B + (XC)_W r_{yc}/r_{xc}. \quad (7\,a)$$

The problem is how to assign a value to k in the general rescaling eqn. (1). Four methods will be considered. In each we assume that regression is linear.

III. Rescaling with Reference to the Criterion

Method 1 may be used if c is known, and the test scores and estimates are to be used together to predict c. The effective predictors are the school test means $\bar{x}$, the deviations $(x-\bar{x})$ about these means, and the corresponding deviations $(z-\bar{z})$ of the original estimates. Denoting these predictors (expressed in standard measure) by m, d and f, and the predicted criterions score by $\hat{c}_s$, we have, $\hat{c}_s = b_1 m + b_2 d + b_3 f$, where b_1, b_2 and b_3 are partial regression coefficients ($b_1 = b_{cm.\,df}$, etc.). We assume that the m's are independent of the d's and the f's and that r_{cd}, r_{cf} and r_{df} respectively are the same for all schools. Then the b's are given by $b_1 = r_{cm}$; $b_2 = (r_{cd} - r_{cf} r_{df})/(1-r_{df}^2)$; $b_3 = (r_{cf} - r_{cd} r_{df})/(1-r_{df}^2)$ (Thomson [12], p. 199).

To find the b's, the following coefficients may be used: $r_{cm} = (XC)_B/\sqrt{(X_B C_T)}$; $r_{cd} = (XC)_W/\sqrt{(C_W C_T)}$; $r_{cf} = (YC)_W/\sqrt{(C_W C_T)}$; $r_{df} = (XY)_W/X_W$. From these b's and r's we can find the multiple correlation of the best weighted composite of these three predictors with the criterion in the usual way (*ibid.*, p. 201).

To obtain weights in working units from these b's, we use: $w_1 = b_1\sqrt{(C_T/X_B)}$; $w_2 = b_2\sqrt{(C_T/X_W)}$; $w_3 = b_3\sqrt{(C_T/X_W)}$: (Snedecor [10], p. 343). These w's

now maximize prediction in the equation $\hat{c} = w_1\bar{x} + w_2(x-\bar{x}) + w_3(y-\bar{y})$. If also we write $\hat{c} = w_2 x + w_3 y$, then $(w_1 - w_2)\bar{x} - w_3\bar{y} = 0$, whence $k_1 = (w_1 - w_2)/w_3$, where k_1 is the constant for Method 1.

Method 2 is applicable if the estimates are to be used alone to predict the criterion. We have $\hat{c}'_s = b_1 m + b_4 f$, where $b_1 = r_{cm}$ and $b_4 = r_{cf}$. The weights in working units are w_1 as before, and $w_4 = r_{cf}\sqrt{(C_T/X_W)}$. Prediction is maximized by using the equation $\hat{c}' = w_1\bar{x} + w_4(y-\bar{y})$, whence $k_2 = w_1/w_4$, where k_2 is the constant for Method 2.

IV. Rescaling without Reference to the Criterion

When the validities of the test scores or estimates are unknown, any method used to rescale the estimates must of necessity be somewhat arbitrary. Two methods appear to be available, differing primarily in the amount of information each employs. Method 3 is based on the 'orthogonal mean square regression' line, i.e. on the principal axis of the ellipsoidal distribution (cf. Greenall [4], and references). Within each school the S.D. and the mean of the estimates are made equal to those of the test scores. This is equivalent to putting $k_3 = 1$ in eqn. (1). Method 4 is based on the ordinary regression line, and for this purpose (after the S.D.'s of the estimates within each school have been reduced to those of the test scores) it uses the average regression within schools of the adjusted estimates on the test scores. This is equivalent to putting $k_4 = r_{xy}$ in eqn. (1): (cf. Gulliksen [5], p. 299, for a somewhat similar use of the ordinary regression line).

V. Comparisons of Rescaling Methods

The four rescaling equations are:

With Method 1, $y_1 = \bar{x}(w_1 - w_2)/w_3 + (z-\bar{z})s_x/s_z$;

with Method 2, $y^2 = \bar{x}w_1/w_4 + (z-\bar{z}s_x/s_z$;

with Method 3, $y_3 = \bar{x} + (z-\bar{z})s_x/s_z$;

with Method 4, $y_4 = r_{xy}\bar{x} + (z-\bar{z})s_x/s_z$.

Methods 1 and 2, while obviously the best, must be ruled out where action has to be taken immediately after the examination in which test and estimates are used. Once the criterion has become available, however, 'follow-back' studies should be undertaken to find the values of k_1 and k_2, and, for Method 1, the best composite of test and adjusted estimates that would maximize prediction. The information thus obtained may be used for similar examinations in the future (see, however, the cautionary note in Lindquist *et al.* [6], p. 692).

The total correlation between the estimates as rescaled by Method 3 and the test scores will be $(X_B + r_{xy}X_W)/X_T$. This correlation is in part spurious, since it is bound to be positive so long as $0 < r_{xy} < 1$, unless of course all schools happen to have the same test mean. The total correlation between estimates rescaled by Method 4 and the test scores is $r_{xy}\sqrt{(X_T)}/\sqrt{(r_{xy}^2 X_B + X_W)}$, which is zero

when $r_{xy} = 0$, and hence not spurious: rescaling by this Method 4 also minimizes the total error of estimate. Hence the total correlation so obtained will be identical with that arrived at by correcting r_{xy} for selection in the bivariate case, as proposed by Gulliksen ([5], p. 135): his formulae (18) and (19) (*ibid*, p. 137–138) for the inferred correlation and overall S.D. are equivalent respectively to the total correlation and the total S.D. of Method 4.

The above considerations, though instructive, are of secondary importance. Here we have to decide, with no criterion to guide us, which of the two methods will be the more useful in practice. Now the total correlation between sets of estimates rescaled by Methods 3 and 4 from the same original estimates is, by eqns. (3 *a*) and (4 *a*), $(r_{xy}X_B+X_W)/\sqrt{\{X_T(r_{xy}^2X_B+X_W)\}}$. If $r_{xy}=0$, this becomes $\sqrt{(X_W/X_T)}$, which is bound to be high, since in practice X_W will not differ greatly from X_T. For likely values of r_{xy} this total correlation will be still higher. Thus little difference is made to the overall order of the children by choosing one method rather than the other. Hence the choice may be made on ground of simplicity in use; and will obviously fall on Method 3.

From eqns. (2 *a*), (5 *a*), (6 *a*) the total validity of the test scores is $\{(XC)_B+(XC)_W\}/\sqrt{(X_T C_T)}$, while, from eqns. (3 *a*), (5 *a*), (7 *a*), that of the estimates rescaled by Method 3 is $\{(XC)_B+(XC)_W r_{yc}/r_{xc}\}/\sqrt{(X_T C_T)}$. Thus the magnitude of the ratio r_{yc}/r_{xc} determines the relative validities of test scores and Method 3 estimates.

The contribution $(XC)_B$ to the total validity of Method 3 estimates is typical of all four. Care must be exercised, therefore, in interpreting these validities. They must be attributed to the composite test-cum-estimates. The phrase 'teachers' estimates'—so frequently used—is somewhat misleading, since it ignores the contribution of the test scores, without which the teachers' original estimates would be of very limited value.

VI. Example

The predicting data here used consist of test scores and teachers' estimates for an almost complete 11+ year-group of 935 children, divided unequally among 18 primary schools in a northern area, where the teachers have had many years' experience with estimates. The test scores are the aggregate $\frac{1}{4}(2I+E+A)$, where I, E and A are 11+ Moray House tests for Verbal Reasoning, English, and Arithmetic respectively, and the education authority adopts Method 3 for purposes of rescaling. The criterion is another Moray House 13+ battery administered two years later.

Table I. Square and Product sums for x, y, and c
18 schools; 935 children. Rescaling by Method 3

Source	(1) $SS(x)$	(2) $SS(y)$	(3) $SS(c)$	(4) $SP(xy)$	(5) $SP(xc)$	(6) $SP(yc)$
Between Schools	20,334	20,334	21,450	20,334	20,223	20,223
Within Schools	150,587	150,587	150,810	135,862	142,646	135,681
Total	170,921	170,921	172,260	156,196	162,869	155,904

TABLE II. INTERCORRELATIONS OF x, y AND c
Rescaling by Method 3

Source	r_{xy}	r_{xc}	r_{yc}
Between Schools	1·0000	0·9683	0·9683
Within Schools	0·9022	0·9466	0·9003
Total	0·9138	0·9492	0·9086

In Tables I and II the estimates (y_3) were the primary teachers' rescaled estimates. The children were subsequently dispersed among different secondary schools. In view of the diverse experiences of the children during the two years before the criterion was obtained, the correlations shown in Table II are remarkably high. The within-school correlation between estimates and criterion scores remains practically the same as that between estimates and test scores. The within-school correlation between test and criterion scores is higher still, as indeed was to be expected in view of the similarity between the two.

The four rescaling methods considered will have in common all the entries in columns 1, 3 and 5 of Table I, and the within-school entries in the other columns.

The b's for Method 1 are found from the data of Table I as follows (see page 195): $r_{cm} = 0{\cdot}3417$; $r_{cd} = 0{\cdot}8857$; $r_{cf} = 0{\cdot}8424$; $r_{df} = 0{\cdot}9022$. Hence $b_1 = 0{\cdot}3417$; $b_2 = 0{\cdot}6754$; $b_3 = 0{\cdot}2331$. From these b's and r's we obtain 0·9546 as the multiple correlation of the best weighted composite of m, d and f with the criterion. Substituting w's for b's in order to get back to the original units we obtain: $w_1 = 0{\cdot}9945$; $w_2 = 0{\cdot}7223$; $w_3 = 0{\cdot}2493$. Hence $k_1 = 1{\cdot}092$, and the rescaling equation is: $y_1 = 1{\cdot}092\bar{x} + (z - \bar{z})s_x/s_z$. Maximum prediction ($r_m = 0{\cdot}9546$) is obtained when the y_1's thus calculated are combined with the x's in the proportion 0·2493 to 0·7223, or approximately 1 to 3. For comparison, the correlation with the criterion of the straight sum ($x + y_3$), obtained from the total correlations in Table II, is 0·9496.

For Method 2, similar calculations lead to $k_2 = 1{\cdot}104$, and hence to the equation $y_2 = 1{\cdot}104\bar{x} + (z - \bar{z})s_x/s_z$. The validity of estimates scaled by Method 2 and used alone is 0·9091.

The results for Method 3 ($k_3 = 1$) are already presented in full in Tables I and II. The rescaling equation is, of course, $y_3 = \bar{x} + (z - \bar{z})s_x/s_z$; and the validity of Method 3 estimates is 0·9086, which should be compared with that of Method 2.

For Method 4, the equation is $y_4 = 0{\cdot}9022\bar{x} + (z - \bar{z})s_x/s_z$; and the insertion of $k_4 = 0{\cdot}9022$ in eqns. (3 *a*) and (7 *a*) gives 0·9071 as the validity of Method 4 estimates.

With large correlations such as these it matters little which method of rescaling is used. In each case k approximates to 1; and the overall order of children will be almost the same for all four methods. Thus, the correlation between y_3 and y_4 sets, calculated as on page 194, is 0·9995.

Method 3 is the customary rescaling procedure; and if the foregoing evidence can be trusted, appears nearly as good as the methods based on known validities.

However, while the above example serves its purpose, the criterion employed is unusual. We must await the results of further empirical studies, carried out along similar lines, but with different criteria, before concluding that the customary method is generally comparable with others which use weights derived from the validities observed in such studies.

References

[1] Burt, C. (1917). *Distribution of Educational Abilities*, p. 14.
[2] Burt, C. (1921). *Mental and Scholastic Tests* (2nd edn.), pp. 439f. and refs.
[3] Burt, C. (1949). Alternative methods of factor analysis. *Brit. J. Psychol.* (*Statist. Sect.*), II, 98–121.
[4] Greenall, P. D. (1949). The concept of equivalent scores in similar tests. *Brit. J. Psychol.* (*Statist. Sect.*), II, 30–40.
[5] Gulliksen, H. (1950). *Theory of Mental Tests.* New York : Wiley.
[6] Lindquist, E. F. (Ed.). (1950). *Educational Measurement.* Washington: American Council of Education.
[7] McClelland, W. (1942). *Selection for Secondary Education.* University of London Press.
[8] McIntosh, D. M., Walker, D. A., MacKay, D. (1949). *The Scaling of Teachers' Marks and Estimates.* Edinburgh : Oliver and Boyd.
[9] Sandon, F. (1956). A regression control chart for use in personnel selection. *Appl. Stat.*, V, 20–31.
[10] Snedecor, G. W. (1950). *Statistical Methods.* Iowa : Iowa State College Press.
[11] Sutcliffe, A., and Canham, J. W. (1945). *Selection for Secondary Education.* Murray.
[12] Thomson, G. H. (1951). *The Factorial Analysis of Human Ability* (5th edn.). University of London Press.
[13] Vernon, P. E. (Ed.). (1957). *Secondary School Selection.* Methuen.
[14] Yates, A., and Pidgeon, D. A. (1957). *Admission to Grammar Schools.* Newnes.

ERRATA

1. In the paper on "Prediction of Scale Values for Combined Stimuli" by H. J. A. Rimoldi (Vol. IX, p. 35) equation 25 should read

$$e = \frac{n\sum(1/V_{jk})}{n'\sum(1/V_j)} :$$

2. In the paper on "Estimation of Factor Analysis under Various Initial Assumptions" by D. N. Lawley (Vol. XI, p. 5, line 26) for

$(M'_{(2)} - L'_{(2)})$ read $(L'_{(2)} - M'_{(2)})$.

Vol. XI Part II The British Journal of Statistical Psychology November 1958

THE CONVERSION OF SCORES TO A UNIFORM SCALE

By M. Howard
University College, London

The original formula for rescaling estimates and test-scores was derived from the three basic assumptions used by Galton and Pearson when deducing their formulae for the correlation coefficient; but it makes no allowance for differences in reliability and validity. However, the various modifications since suggested appear to rest on highly speculative arguments, and the improvements so obtained are seldom large enough to be statistically significant. On the basis of actual trials with tests and estimates of average validity, assessments of the predictive accuracy to be expected with the original formula have been computed. The accuracy diminishes progressively with the time-interval involved, but remains high enough to justify its use for the practical purposes of the educationist.

I. Origin of the Problem

In several recent publications by myself and others in this journal and elsewhere (e.g. [26], [27]), reference has been made to the use of teachers' estimates of pupils' abilities to adjust or correct the scores obtained with tests of general intelligence or the like and of test scores to adjust or correct the varying standards of teachers; and a number of critics have asked for fuller details regarding the procedure employed. In particular it has been suggested that the method usually adopted, both for theoretical inquiries and for practical purposes, in London schools makes no allowance for differences in reliability or validity, and thus differs somewhat from those used by Thomson, McClelland, and Pilliner in their work in Scottish schools, and discussed more fully by Mr. Pilliner in the foregoing paper[30][1]. The object of this note therefore is to examine a little more closely the basis and the limitations of the methods we have used.

As Mr. Pilliner has pointed out, recent investigators seem for the most part to have adopted a rule-of-thumb procedure without considering the underlying assumptions that might justify, or fail to justify, the methods employed. He himself does not carry the history of the subject further back than McClelland's work on *Selection for Secondary Education* (1942) in Scotland. One may perhaps hazard the conjecture that these later writers omitted to offer formal proof for their formulae because they assumed that readers would be sufficiently familiar with earlier discussions of the problem. Since those discussions now seem to be largely forgotten or difficult to procure, it may be worth while to trace quite briefly the history of the subject back to its early origin, and incidentally to see how far the earlier arguments need to be corrected or modified in the light of further experience.

[1] I am grateful to both the author and the editor for the opportunity of seeing Mr. Pilliner's paper in advance, and have greatly profited by his analysis of the difficulties involved.

In psychological and educational work the statistical problem seems first to have arisen in connection with the validation of intelligence tests when they were being put forward for practical use. In this country it has been the custom from the very outset (Burt [5]) to validate such tests by applying them to *complete age-groups* (not to grades or classes as in later American researches) and to adopt as the criterion an independent estimate of the children's ability, obtained from experienced teachers who know them individually. Pupils in a given age-group, however, are usually spread over several different classes; and, to secure reasonably large samples, it was further necessary to include pupils from several different schools. In such a case there will be no one teacher who knows every child in the entire group; and different teachers have different standards and different methods of spreading their marks. Hence, to render their estimates reasonably comparable it became essential to re-scale them so that they could be placed on a single scale that would be the same for all[1].

In this country few of the earlier educational psychologists put complete trust in the unchecked results obtained from tests alone. Whether the aim was to detect cases of certifiable ' mental deficiency ' suitable for transference to special m.d. schools or to select children for free places at a secondary school, the test results were almost invariably supplemented, especially in borderline cases, by estimates from teachers who knew the children personally. The practical problems for which some method of rescaling was most urgently needed, and was most frequently used (at least till the date on which the recent Education Act came into force), arose in connection with the annual junior county scholarship examinations, arranged by the London County Council, for pupils of 11 plus.

II. Basic Assumptions

In solving these various problems, both theoretical and practical, the simplest and most obvious assumptions to adopt were those originally suggested by Galton [2] when developing his method of measuring correlations, and subsequently taken over by Karl Pearson as the basis of his product-moment formula. Galton and Pearson had in mind chiefly physical measurements. The first to apply them to psychological assessments seems to have been Burt, whose formulation I have adopted, with slight modifications in the wording to bring the arguments into line with modern terminology. He states them in postulational form.

[1] According to Greenall [20], " in the field of psychology and education the problem was first systematically discussed shortly before the first world war, by the British Association Committee appointed to inquire into ' Factors in Education ' (1911) with J. A. Green and C. Burt as secretaries ". Their formulae were expressed in terms of the probable error, and used in connection with the validation of various standardized mental and scholastic tests. At that time, however, the popularity of ' intelligence quotients ', ' achievement quotients', and other devices for expressing estimates as mental ratios, tended to obscure the merits of statistical units of a more precise and technical type; and the first systematic use of the standard deviation as a universal unit appears to be that described by Burt in his various L.C.C. Reports [7, 8, 9].

" Galton's procedure rests on three basic assumptions, which may be expressed in general terms as follows: (i) average scores are comparable ; (ii) average deviations are comparable ; (iii) the same proportions that hold between average deviations hold for all equivalent deviations from their respective averages "[1]. As a measure of the ' average ' score and ' average ' deviation Galton took the median and the quartile deviation. Pearson proposed instead the arithmetic mean and the root mean square (or ' standard ') deviation as being more reliable. The rescaled scores, expressed in terms of the mean as zero and the standard deviation as unit, are said to be in ' standard measure '. Hence the foregoing assumptions are summed up by Burt in a single phrase : " equal standard measures are comparable "[2].

As applied to the problems of the educationist these basic requirements may be reformulated more specifically as follows.

(i) A mean score calculated for a particular school (or other subgroup) from the teacher's estimates should indicate the same degree of ability as the mean score for the same subgroup obtained with a formal test applicable to all the subgroups.

(ii) An estimate for an individual pupil which deviates from the mean estimate for the subgroup by one standard deviation (as calculated from the estimates for that subgroup) should indicate the same degree of capacity as a test-score which deviates in the same direction by one standard deviation (as calculated from the test-scores for the same subgroup).

(iii) If they bear the same proportion to the standard deviations of (*a*) the estimates and (*b*) the test scores respectively, then the deviations of two measures, whether obtained from estimates or from test-scores, should indicate the same degree of ability.

III. Basic Formula

With these assumptions a simple working formula may be deduced as follows (cf. Burt [7], [8], and refs.: instead of his more general terms ' subgroups ', ' subtests ', and ' criterion ', it will probably be clearer to refer specifically to ' schools ', ' estimates ', and ' the overall test ' respectively).

Let y_{ik} denote the mark obtained by the ith individual ($i = 1, \ldots n_k$) in the kth school ($k = 1, \ldots N$) with the kth set of estimates, i.e. those available for the kth school *only*. Let X_{ik} denote the mark obtained by the same individual with the ' universal ' test (or other general criterion) available for *all* schools. And

[1] The term ' average ' is here used in the broad sense defined by Yule ([6], pp. 108f. and p.134), and the term ' comparable ' in the sense defined by Kelley ([11], p. 114).

[2] Galton [2] first developed the method in order to ' co-relate ' measurements of bodily height obtained either for parents and their children or for male and female adults. The clearest formulation of the principles implied is perhaps that given by T. L. Kelley, who worked under Pearson, and re-states Galton's method in the form which Pearson gave it (cf. Kelley, [11], pp. 115f.). Of the three basic assumptions set out above the last always calls for preliminary verification on the basis of empirical data wherever this is at all practicable (a point on which Kelley also insists).

let Y_{ik} denote the theoretical mark that would be obtained by the same individual on the basis of the kth set of estimates, after converting the mean and the dispersion of those estimates to terms of units that will hold for *all* the individuals in *all* the schools. (Capital letters, it will be noted, refer to measurements for the entire group; small letters to measurements for the smaller groups.)

The data available will as a rule consist of the values actually obtained for X_{ik} and y_{ik} for the $n_1 + \ldots + n_N$ children in the N schools: it is required to find Y_{ik} for any child in any school. For simplicity let us confine the argument to one of the schools only. We can then drop the subscript denoting it (k). Accordingly, in virtue of the foregoing postulates, we assume

$$\overline{Y} = \overline{X}, \tag{1}$$

$$\sigma_Y = \sigma_X, \tag{2}$$

and

$$\frac{Y_i - \overline{Y}}{\sigma_Y} = \frac{y_i - \overline{y}}{\sigma_y}. \tag{3}$$

Then, on substituting in eqn. (3) the values given by eqns. (1) and (2), we obtain at once

$$\frac{Y_i - \overline{X}}{\sigma_X} = \frac{y_i - \overline{y}}{\sigma_y},$$

i.e.

$$Y_i = (y_i - \overline{y})\sigma_X/\sigma_y + \overline{X}. \tag{4}$$

This is the formula in common use. Thomson's proposals for dealing with the same and kindred difficulties are based on a similar equation, reached by slightly different arguments[1].

Equation (4) has the form of a ' freedom equation ' for a straight line relating variations in Y to variations in y, with $\overline{X}$ as the intercept on the axis of Y, and $\tan^{-1}(\sigma_X/\sigma_y)$ as the angle of inclination to the axis of y. For this reason presumably Otis [10] refers to it as the ' relation line '. Greenall [20] prefers to call it the ' equivalence line '. The argument and the resulting equation may obviously be generalized to cover the cases in which a multiplicity of criteria, or a multiplicity of estimates, or both, are involved.

As Burt and others have pointed out, a theoretical justification for preferring this equation and this line to others that have been proposed may be reached by following the argument adopted by Karl Pearson [3] in deducing his method of ' principal axes '. The ' equivalence line ' is in fact identical with that which represents the ' first ' general factor as obtained by ' weighted summation ' (i.e. by Pearson's application of ' least squares '). The clearest and most general exposition of Pearson's procedure is given by Cramér, who terms the line the ' orthogonal mean square regression line '. (See [18], p. 284 for the

[1] Thomson [12], esp. pp. 400f., and [17], esp. p. 78f. Pilliner's method of deriving his four main formulae is based on two main assumptions, which may perhaps be regarded as a more general version of those proposed by Galton and Pearson, and adopted by Burt. Except for a difference in notation, eqn. (4) in the text above is identical with Pilliner's equation for his ' Method 3 '.

bivariate case, and p. 309 for the multivariate case ; also [21], pp. 106–7. Still greater precision can be secured by the use of so-called ' probit methods ': cf. Burt [23], esp. pp. 113f.)

However, when used to express the relation between two variates, Pearson's procedure assumes, as Yule has pointed out, that both are subject to error in equal degrees. When they are subject to error in unequal degrees, then each should be weighted in direct proportion to its precision, as the traditional method of least squares requires[1]. In the early days of mental testing it was in fact commonly found that the reliability and validity of the teachers' estimates were much higher than those of the tests, but that they varied far more widely from one teacher to another. In such cases it was customary to allot differential weights accordingly, and calculate the ' equivalence line ' by the method adopted in the modified version of weighted summation ([21], p. 113). This is the device I myself have adopted when there was a fairly large difference in the reliabilities.

The formula we have given (eqn. (4) above) is the earliest of its kind. Several other procedures have since been developed. But all of them, it would seem, including those described by Greenall and by Pilliner, can be summed up by generalizing eqn. (4), and rewriting it in the form

$$y_i = w_y(y_i - \bar{y})\sigma_X/\sigma_y + w_X\overline{X}. \qquad (5)$$

Different assumptions and different models then lead to slightly different values for w_y and w_X.

It is natural to inquire (i) which of them yields the best procedure on theoretical grounds and (ii) which is the most appropriate for practical use. Where, I think, I should differ from both Pilliner and Greenall, is in the answer to these questions: they apparently assume that there is only *one* best method, suitable for all types of problem. In my view each has its special merits and its special limitations : so that one may be suited to one type of problem, others to other types. Unfortunately, nearly all the more refined procedures involve a process of successive approximation, and thus greatly increase the labour of calculation for investigators who have no access to electronic computers. However, with data of the kind that are nowadays available differences in reliability are usually slight. And we ourselves have found that the improvements obtained by the more elaborate equations are, as a rule, statistically nonsignificant, and therefore scarcely worth the additional work entailed.

[1] From a theoretical standpoint the best discussion of this problem is that contained in Wald's paper [16]. The convenient term ' orthogonal ' regression (to denote " the regression obtained by minimizing the square-sum of the deviations measured perpendicularly to the regression ") seems due originally to Frisch ([13], p. 65). As he points out, " if this is to be used in a case where the units of measurement are conventional, it should be applied to the *normalized* variates " (i.e., variates in standard measure): otherwise it will *not* be invariant for a change in the units of measurement (cf. *op. cit.*, ch.X, on ' The Interpretation of the Empirical Regressions—Elementary, Orthogonal, Duo-orthogonal, and Diagonal ', which discusses the relations between the various regressions and the principal components or ' factors '). This warning seems to furnish an answer to Greenall's criticisms of the ' line of closest fit '.

IV. Illustrative Results

Mr. Pilliner very rightly contends that the main test for the validity of the various procedures, and for the assumptions on which they rest, must in the last resort turn on the actual results obtained; and, since the conclusion derived from a single sample only cannot by itself carry final conviction, he appeals for similar analyses based on independent data. Indeed, quite apart from theoretical considerations, the amount of predictive accuracy that is actually achieved by using such devices is itself a matter of considerable practical interest.

The samples which I have myself analysed are smaller than his. And the only follow-up data that are available for a complete age-group relate to a group of just over 400 children, whose abilities were tested and estimated at the age of 10–11, and of whom 368 were regularly re-assessed year by year to the age of 13–14 or later. We also have data for 473 children tested during the junior county scholarship examinations. This latter group includes no backward or mentally defective children, and indeed comparatively few below average intelligence. As a result the correlations are smaller, but are comparable amongst themselves.

In all these cases the teacher was allowed to submit his marks in whatever form he wished; and they were then reduced (by means of a general graph) to terms of the teacher's own mean and standard deviation. If he preferred to rank his pupils in order of merit, the rectangular scale of ranks was converted to a normalized scale in standard measure by the aid of the usual percentile tables: the procedure adopted was that originally suggested by Galton [1, 4], and is fully discussed in most of the textbooks on psychological measurement (e.g. [15], ch. VIII and conversion table p. 249: cf. [8], pp. 47f.). The reliability coefficients were based on parallel tests or parallel estimates, and were procured from somewhat smaller groups. In some instances it was possible to check the figures so obtained by calculating reliability coefficients for the entire group by the variance method ([25] and refs.): the coefficients thus obtained were somewhat higher, but their relative magnitude remains much the same. The various correlations obtained for the validity and the reliability are shown in Table I.

It will be seen that the improvement in the validity coefficients gained by differential weighting is too small to be statistically significant. Our figures for validity are not so high as those reported by Mr. Pilliner, but decidedly higher than the values suggested by the British Psychological Society's report on *Secondary School Selection* [28]: the authors of the report apparently accept as 'the most satisfactory study' that of Husen [22], who obtained a correlation of 0·72 with Swedish children. The only English investigation which they report is that of Burt [24], who found "a correlation of 0·84 for 782 cases tested at the ages of 10 to 12 and again 8 or more years later: the tests applied on the two occasions" (they add) "were rather closely parallel, and their reliability was increased by supplementary individual testing of cases whose original results were queried by teachers: thus this correlation is higher than those obtained in any other study except perhaps Honzik's [19], which furnished a correlation of

0·83 obtained after intervals of about 6 to 18 years''. In point of fact, however, figures almost as high have been reported even by American investigators: Byrns and Henmon [14], for example, found a correlation of 0·81 for College students tested 8 to 10 years previously: allowing for the restricted range this is quite compatible with a figure of 0·85 for a complete age-group. On the basis of data for children tested at 11+ and followed up to 13–14 years, Yates and Pidgeon [29] claim ''validity coefficients of 0·90 or slightly more''.

TABLE I. RELIABILITY AND VALIDITY OF TEACHERS' ESTIMATES AND TEST SCORES

Method of Assessment	1. Complete Age-Group		2. Candidates for the J.C.S. Examination	
	Reliability	Validity	Reliability	Validity
Tests:				
All Schools	0·92	0·76	0·91	0·71
Average within Schools	0·87	0·69	0·75	0·68
Teachers' Estimates:				
Uncorrected:				
All Schools	0·55	0·49	0·46	0·52
Average within Schools	0·94	0·78	0·93	0·76
Teachers' Estimates:				
Corrected and Weighted:				
All Schools	0·97	0·88	0·85	0·74
Corrected but Unweighted:				
All Schools	0·93	0·86	0·81	0·72

It should be noted that the correlations commonly quoted from American writers refer solely to test-results taken just as they are, and to groups which rarely represent complete age-groups. In the view of most competent writers in this country it is essential to deal with samples selected in such a way that complete age-groups are represented: otherwise it is almost impossible to allow for variations in the range covered by the individuals tested. Even then, it must be remembered, the results will only hold good for the populations with which the investigator or the local authority concerned is dealing.

V. CONCLUSIONS

So far as theoretical considerations are concerned, it may, I think, be legitimately concluded from the foregoing figures that the original rescaling equation in its simple form is sufficiently accurate for the majority of purposes. The introduction of further coefficients, based on some hypothetical scheme of differential weighting, rests on highly speculative arguments; requires accurate data which are seldom available; and produces changes in the final figures which are too small to be statistically significant.

So far as the practical results are concerned, it would seem that the amount of predictive accuracy that we can expect with the method of rescaling here

recommended lies about midway between the more extreme estimates suggested by the recent writers. The figures suggested by Pidgeon and Yates seem a little too optimistic; those given by the Working Party of the British Psychological Society somewhat too pessimistic. So far as we may judge from our experience in London it appears that the correlations in question tend as a rule to drop slowly from about 0·95 (for parallel tests applied at the same time) to 0·90 after 1 year, 0·875 after 2 years, 0·85 after 3 years, and so on, by almost regular annual increments.

References

[1] Galton, F. (1883). *Inquiries into Human Faculty and its Development.* Macmillan.

[2] Galton, F. (1888). Correlations and their measurement. *Proc. Roy. Soc.*, XLV, 136–145.

[3] Pearson, K. (1901). On lines and planes of closest fit to points in space. *Phil. Mag.*, II, 559–72.

[4] Galton, F. (1907). Grades and deviates. *Biometrika*, V, 67–81.

[5] Burt, C. (1909). Experimental tests of general intelligence. *Brit. J. Psychol.*, III, 94–176.

[6] Yule, G. U. (1910). *Introduction to the Theory of Statistics.* (2nd ed., 1912; 11th ed. with M. Kendall, 1937). Griffin.

[7] Burt, C. (1914). *Annual Report of the Psychologist to the London County Council.* P. S. King.

[8] Burt, C. (1917). *Distribution and Relations of Educational Abilities.* P. S. King.

[9] Burt, C. (1921). *Mental and Scholastic Tests.* P. S. King.

[10] Otis, A. (1922). A method of finding the correspondence between scores in two tests. *J. educ. Psychol.*, XIII, 529–47.

[11] Kelley, T. L. (1923). *Statistical Method.* New York: Macmillan.

[12] Thomson, G. H. (1928). The mental age concept and the standardization of group tests. *Psychol. Rev.*, 398–413.

[13] Frisch, R. (1934). *Statistical Confluence Analysis by Means of Complete Regression Systems.* (Reprinted from the *Nordic Statistical Journal*, V, 5). Oslo: Universitets Okonomiske Institutt.

[14] Byrns, R. and Henmon, V. A. C. (1935). Long range prediction of college achievement. *School and Society*, XLI, 877–880.

[15] Guilford, J. P. (1936). *Psychometric Methods.* New York: McGraw Hill.

[16] Wald, A. (1940). The fitting of straight lines when both variables are subject to error. *Ann. Math. Stats.*, XI, 248–300.

[17] Thomson, G. H. (1945). The distribution of intelligence among college students. *Brit. J. educ. Psychol.*, XV, 76–79.

[18] Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton: University Press.

[19] Honzik, M. P. *et al.* (1948). The stability of mental test performances. *J. exp. Educ.*, XVII, 309–24.

[20] Greenall, P. D. (1949). The concept of equivalent scores in similar tests. *Brit. J. statist. Psychol.*, II, 30–40.

[21] Burt, C. (1949). Alternative methods of factor analysis and their relations to Pearson's method of ' principal axes '. *Brit. J. statist. Psychol.*, II, 98–121.

[22] Hüsen, T. (1950). *Test resultatens Prognosvärde.* Stockholm: Gebers Vörlag.

[23] Burt, C. (1951). Test-construction and the scaling of items. *Brit. J. Psychol. Statist. Sect.*, IV, 95–129.

[24] Burt, C. (1954). Age, ability, and aptitude. *Univ. Lond. Inst. Educ. Stud. Educ.*, vi. Evans.
[25] Burt, C. (1955). Test reliability estimates by analysis of variance. *Brit. J. statist. Psychol.*, VIII, 103–118.
[26] Burt, C. (1955). Evidence for the concept of intelligence. *Brit. J. educ. Psychol.*, XXV, 158–177.
[27] Burt, C. and Howard, M. (1956). The multifactorial theory of inheritance and its application to intelligence. *Brit. J. Statist. Psychol.*, IX, 95–131.
[28] Vernon, P. E., (ed.) (1957). *Secondary School Selection.* Methuen.
[29] Yates, A. and Pidgeon, D. A. (1957). *Admission to Grammar Schools.* Newnes.
[30] Pilliner, A. E. G. (1958). The rescaling of teacher's estimates. *Brit. J. statist. Psychol.*, XI, 191–197.

NOTES AND CORRESPONDENCE

HEREDITY AND BEHAVIOUR

In a recent paper in this *Journal* (X, p. 62) Burt and Howard conclude by saying that, to reach satisfactory evidence as to the inheritance of mental characteristics in man, " mere statistical inquiries are not enough : an adequate understanding of the processes involved can be gained only by carefully planned experiments on lowlier animals ". In this country most experimental studies of the genetics of behaviour (e.g. those of McDougall at Oxford, Miss Smith at Cambridge, Miss Hazlitt at Bedford College, Jonckheere and Audley at University College) have been carried out with small mammals such as the rat. With these, however, each research has to be prolonged for many years before adequate results can be secured. The purpose of this preliminary note is to argue, on the basis of trials with several different species, that the most suitable mode of approaching the whole problem is to start with the organism which has figured so prominently in genetic studies of physical characteristics, namely, the small fruit fly, *Drosophila*. We are indebted to Professor Burt for this suggestion and for advice as to the general plan of the experiment, to Dr. H. Kalmus (also of University College) for helpful references to the literature[1], and to Dr. R. F. Morton for a supply of organisms. In the conduct of the experiments and the statistical analysis of the data valuable assistance has been given by Miss L. Dutton and Miss G. E. Gibbs.

The plan of the experiment broadly followed the procedure outlined in the paper cited (p. 61). (i) The first step is obviously to decide on and define as precisely as possible the population to be studied. The advantage of choosing *Drosophila melanogaster* is that this species can be bred in large numbers with little trouble and expense. A single generation, from egg to egg, is completed in about 10 days ; hundreds of offspring are produced by each pair. And the problem is greatly simplified owing to the fact that this species possesses only four pairs of chromosomes, which have already been fairly thoroughly mapped, whereas the rat has as many as 42. Our initial sample was drawn from a stock of unselected flies with a history of random mating, and consisted of 100 individuals. (ii) Having defined the initial population, the next step is to define the ability or trait to be observed. Our ultimate aim was to check the hypothesis of an ' innate general cybernetic efficiency '. We started with simple phototropic and (negative) geotropic reactions (i.e. flying in the direction of light and flying upwards against the force of gravity). These can be defined more exactly in operational terms by describing the procedures used for measuring them. We have been less successful in discovering other practicable manifestations of ' cybernetic efficiency '. (iii) As a result of previous work with correlational techniques, we now fully agree that, for both experimental and statistical purposes, it is " more effective to proceed by determining the differences between individuals than by determining resemblances within a group, as has been the common practice with psychologists hitherto ". Accordingly, the third step was to ascertain the distribution of these differences for the traits defined. The method we have adopted is what Hirsch and Tryon call ' mass screening '. This aims at constructing a linear scale by a succession of binary or dichotomous classifications (cf. this *Journal*, IV, p. 196). We modified the procedure there described by adding the positive and negative ' factor measurements ', thus obtaining, with n dichotomies, a scale of $n+1$ intervals, not 2^n.

[1] Of the numerous investigations hitherto published, hardly any deal with individual differences in behaviour as distinct from physical characteristics. By far the most useful has been that by J. Hirsch and R. C. Tryon describing investigations carried out at the University of California (' Mass Screening and Reliable Individual Measurement in the Experimental Behaviour Genetics of Lower Organisms ', *Psychol. Bull.*, 1956, LIII, pp. 402–410).

To measure the geotropic reaction we begin by placing all the individuals at the bottom of a wide-mouthed bottle, with a second inverted and placed above, mouth to mouth. After 10 to 15 seconds, about half will have flown into the upper bottle, thus ' passing ' the first ' test '. A card is then slipped between the two bottles. The procedure is repeated with the group in each bottle, and so on, for 9 consecutive tests. We thus obtain 10 different subgroups, varying in speed of reaction. A reliability coefficient can be calculated by means of the formulae described in an earlier number of the *Journal* (VIII, 103–118 ; cf. also Hirsch and Tryon, *loc. cit.*, pp. 407–9). In our own experiments it averaged 0·91. (iv) The extreme groups thus selected were then used for breeding; and the means and variance determined afresh for the new generations. With this organism it is also possible to base selection on particular chromosomes, or even sections of them, as has so often been done when studying the genetics of its physical characteristics. The final selection can be rendered so thorough that the members of the sub-samples eventually obtained are almost as similar as identical twins. If we may trust our own preliminary trials, it would seem that with the fruit fly 8 or 10 months will suffice to secure a relatively homogeneous sample which could only be achieved with rats in an investigation lasting 8 or 10 years. (v) The final step is to compare means and variances and determine the statistical significance of the differences.

The average mark for the offspring of those who had achieved a mark of 1 or 2 on a 10 point scale was 3·1 ; whereas for the offspring of those who had achieved a mark of 8 or 9 it was 7·2—the difference being fully significant at the 5 per cent level. Evidently then the tendency underlying this particular reaction is largely influenced by heredity. And generally the results so far obtained fully confirm the suggestion that " many of the experimental designs used for studying the genetics of physical characteristics in *Drosophila* can be adapted in studying the genetics of behavioural differences ". It is always possible to estimate the correlations from the variances. Hence this mode of approach incidentally enables us to ascertain how far the results of a mere correlational analysis would agree with those of an experimental analysis, and at the same time may throw considerable light on the genetic and post-genetic mechanisms at work to produce the observable phenotypic differences.

If possible it would be desirable to estimate behaviour for several different types of activity. In the study of human psychogenetics, the question chiefly to the fore at the moment is the inheritability of the general cybernetic factor of ' intelligence '. A similar issue has arisen in the studies of lowlier mammals such as rats and mice. Professor Krech, for example, seems to incline to the hypothesis of a single general factor (similar to that described by Spearman) on the ground that the central nervous system would have much the same general efficiency throughout and tends to operate as a unitary whole. Professor Tryon, on the other hand, apparently inclines to a hypothesis similar to that put forward by Thurstone: he holds that behaviour is controlled by a number of independent specific factors, on the ground that inheritance is determined by separate unit-factors determined by independent genes. The multifactorial theory put forward by Burt and Howard (this *Journal*, IX, pp. 100f.) suggests a possible reconciliation; for, if the general efficiency of the central nervous system was determined by a *number* of independent factors, the resulting variations in behaviour would tend to show a graded distribution, more or less approximating to the normal. For this problem it would be desirable to measure at least half-a-dozen distinct characteristics, e.g. rate of development[1], speed of reaction to different stimuli, and, if possible, certain forms of conditioning or learning. So far we have failed to discover activities or tests which would satisfactorily fulfil this programme. The ingenuity of other investigators, however, may help to solve these more ambitious problems.

At the outset of our investigations our original intention was primarily to arouse the interest of sceptical school teachers in the influence of genetics on behaviour generally:

[1] Cf. J. M. and S. M. Smith, 'Genetics and Cytology of *Drosophila* ' *Subobscura*, VIII, Heterozygosity and Rate of Development, *J. Genet.*, L 11, 1954, pp. 152.

we found that experiments along these lines, with demonstrations, led to a clear appreciation of the issues involved and a keener understanding of the conclusions that might be reached. As the work progressed, however, it seemed likely that the method might even serve to make a positive contribution to serious research. We hope to report the detailed results in a later issue of this *Journal*. Meanwhile, we believe it might be worth while drawing the attention of readers to the remarkable ease with which such experiments can be carried out and the obvious value of studies along these lines.

J. L. PHILLIPS.

FACTORS OF NEUROTICISM

To the Editor, British Journal of Statistical Psychology.

SIR,—In an earlier issue of this *Journal* and in one or two more recent papers, Dr. Gregory has put forward certain novel views on the nature of neuroticism, based, it would seem, on statistical investigations. His aim, if I understand him rightly, is to suggest an ingenious hypothesis which he regards an alternative to, and an improvement on, the ' dimensional hypothesis ' proposed by Dr. Eysenck. Unfortunately, however, his main conclusions, as he states them, seem to be mutually inconsistent or at least highly paradoxical.

In one place, for example, he argues that " the difference between a man who is, and a man who is not, suffering from some neurotic or psychotic disorder, is a difference, not of degree but of kind, and therefore to be expressed in terms not of quantitative variables, but of qualitative attributes " (this *Journal*, V, p. 210) ; yet on the following page he declares that " the factor of general emotionality " (where individual differences are admittedly expressed by " differences of degree " on a continuous scale) " is virtually identical with the factor of neuroticism " (*ibid.*, V, p. 211, cf. also VI, p. 51). Again he says, " psychodynamics treats the behaviour of a human being largely on the analogy of the behaviour of a physical system acting under the influence of internal and external forces. The forces may be in conflict : and then, when internal, they produce friction ; when external, they set up stress or strain. According as they vary in strength and direction, the individual may be in a state either of stable equilibrium or of instability. Different individuals show different degrees of resistance of strain and stress ". These forces, we are told, " can, in theory at least, be measured ". So far, therefore, we are intended to take the analogy quite literally. But suddenly we find ourselves in a context of vague and mixed metaphors : we read of ' streamlined ' behaviour, of the ' flutter ' of the neurotic, the ' emotional turbulence ' of the young delinquent, and of ' oscillations in a goal-seeking process '—each of which somehow or other may " produce a new and qualitatively different state, and point to an impending collapse ". Finally, to render the confusion worse confounded, Dr. Gregory ends with " a reminder that organisms are something more than mere physico-chemical mechanisms : they are alive ".

All this is made the basis for an attack upon the " dimensional interpretation of neuroticism ". Thus the main question raised by these somewhat confusing arguments seems to be this : is the difference between normal and neurotic individuals merely a difference of quantity or degree (as Dr. Eysenck and others maintain) or is there some qualitative difference between the two ? If Dr. Gregory wants us to reject the simpler and well-established dimensional hypothesis in favour of the latter, will he tell us what precisely is his empirical evidence ?

CATHERINE LAMBERT,

THE FACTORS OF NEUROTICISM : A REPLY

The chief empirical evidence for the view which Dr. Lambert criticizes was derived from a set of assessments for the behaviour of about 420 school children, both normal and neurotic. Factor analysis revealed a large general factor (which, it was suggested, might be termed ' stability ') quite independent of the factors for ' general emotionality ', and for repression *versus* non-repression. The factor measurements exhibited a bimodal distribution ; and the distinction between the two subgroups corresponded (apart from a 3 per cent overlap) with the distinction between normal and neurotic children as independently diagnosed. My contention was that this sharp division between the two groups implied a qualitative difference.

The ' hypothesis ' to which Dr. Lambert refers was not put forward as an " an alternative to the dimensional hypothesis proposed by Dr. Eysenck ", but was based on an earlier theory of Professor Burt's ; and most of the statements which she cites were in fact quotations or paraphrases from his papers or lectures : (cf. ' The Unstable Child ', *Child Study*, X, pp. 61–78, and *The Subnormal Mind*, 3rd., pp. viii and 195f. and refs.).

Briefly the argument was that the behaviour to be studied should not be interpreted in terms of a simple stimulus-response process or of conditioned or unconditioned reflexes, but of a ' flow ' of nerve currents from the afferent to the efferent side, so numerous and compact as to be capable of being treated for descriptive purposes as a single stream. The changing characteristics of the ' flow ' of this stream express themselves by the changing characteristics in the ' flow ' of behaviour. On this basis, we are told, " it should be possible to formulate a hypothetical model derived from the theory of hydrodynamics ". Thus neurotic ' breakdown ' is not to be conceived (as is commonly the case) on the analogy of " a thin plank that cracks and gives way because the stress imposed by a superincumbent weight is greater then the structure can bear, or a rope that is stretched by excessive tension until it breaks "—a model which is grossly oversimplified, but rather as " a loss of stability produced when the intensity of certain specifiable factors or forces is too great for the directional drive to be preserved by the ordinary mechanisms of ' feedback ' ". The words ' flutter ' and ' turbulence ' were intended, not as ' vague mataphors ', but as technical terms having precisely the same meaning that they are given in the theory of liquids, in aerodynamics, and in the study of modern rocketry[1].

The fact that a *quantitative* increase in one variable (e.g. pressure, velocity, or temperature) may, at certain critical points, produce new ' states ' exhibiting entirely different *qualitative* features is not so paradoxical as Dr. Lambert supposes. In physics the most familiar example is the way quantitative changes in the mean kinetic energy of molecular motion produce qualitative changes in the observable properties of matter in bulk—e.g. in the transformation of substances from the solid state to the liquid, and from the liquid to the gaseous.

In my view the difference between a state of emotional equilibrium or stability—a ' steady ' state—and a state of emotional instability is essentially qualitative. It appeared to correspond, in the sample of the child population described above, with the distinction between the ' normal ' and the ' pathological ', and a ' pathological ' condition similarly implies a qualitative difference. But most of the underlying factors responsible for the

[1] When a fluid (liquid or gas) flows with a low velocity through a tube having a smooth, uniform, and narrow bore, the flow is said to be ' streamline', i.e. the velocity contains no components transverse to the main line of the stream : the speed (or acceleration) is uniform, and the direction steadily maintained. When the bore of the channel is enlarged (as in a wind tunnel) or becomes irregular, then, as the pressure at the inflow is increased, a critical point is reached at which the motion becomes ' turbulent ', i.e. eddies and minor currents are formed, and the flow becomes erratic both in rate and in direction. L. F. Richardson in his meteorological researches studied similar phenomena in the atmosphere ; and some of the formulae used in analysing ' emotional turbulence ' were originally suggested by him.

difference (general emotionality, general intelligence, environmental stress, social repression, etc.) exhibit measurable differences in amount or intensity only.

My reasons for preferring this dynamic model to the simpler ' dimensional hypothesis ' that Dr. Lambert favours are not confined to the results obtained by statistical analysis. They rest quite as much on clinical observation. The dimensional hypothesis merely enables us to make a rough threefold classification of individual cases. The model proposed by Burt, particularly his theory of ' emotional turbulence ', enable one to offer an intelligible interpretation for numerous peculiarities in the behavioural symptoms and the causal histories of individual cases, and at times may even suggest appropriate lines of treatment. Above all it provides a clearly defined set of concepts in place of the vague, semi-popular language usually adopted in books and papers on child psychiatry. Except for a few empirical studies, Burt's accounts are for the most part somewhat fragmentary and non-technical. My aim has been to re-state the hypothesis in a set of formal propositions, deduce corollaries, and test them by actual data. The detailed results will, I hope, be published in a forthcoming issue of this *Journal*. Here I need only say that, so far as they have been analysed at present, they indicate that the model supplies a much better working basis than any other equally precise hypothesis. Nevertheless, it still seems to leave a good deal unexplained. Hence my final reservation.

W. L. Gregory.

BOOK REVIEWS

Secondary Modern Schools. By H. C. Dent. Routledge & Kegan Paul, 1958. Pp. xvi + 207. 15*s*.

According to the latest figures available the proportion of children in England and Wales transferred at the age of 11+ to Secondary Modern Schools and to Grammar or Technical Schools is 75·4 per cent and 24·6 per cent respectively. On the conventional scale this would imply an average borderline of about 110 I.Q. The Secondary Modern School thus has to cope with an unusually wide range of ability whose upper and lower limits vary considerably in different areas. It is therefore not surprising to find that, during the past few years or so, schools of this new type have broken up into a number of widely divergent subtypes: Professor Dale distinguishes half-a-dozen, each of which he describes in detail on the basis of personal visits to schools all over the country.

One point stands out very clearly. The common complaint that " the Secondary Modern School is merely an old-fashioned Elementary School masquerading under a new name " is quite unjustified. There is (as he puts it) an " emerging pattern of education which genuinely merits the title ' Secondary ' ". The vast majority of parents, he finds, now seem quite content with the education their children are receiving in the Secondary Modern School. Naturally those who had hoped that their children would benefit by the social and economic advantages that are supposed almost automatically to follow selection for a Grammar School experience a painful disappointment when their children are rejected. Many are therefore tempted to align themselves with the critics whose favourite target is the 11+ examination. But, as Professor Dent observes, the real object of dread and dislike is not so much the method of selection but the sheer fact of selection. Nevertheless, he believes, the discontented have been " vocal out of all proportion to their number "; and their grievances have been magnified both by politicians and by the press. As a result several education committees have aimed at a wider development of the ' Comprehensive ' school, but more often, he believes, " for political than for educational reasons ".

In his view, however, it is the organization of the Secondary Modern School rather than the mode of recruitment that really calls for critical study. The total range of pupils

with which the Secondary Modern School has to cope may, he thinks, be conveniently subdivided into three main sections. There are first the brightest of all—the borderline cases who have just missed allocation to a grammar or technical school, and therefore overlap with the duller pupils who have successfully scraped through the various tests: in the old days these would have been largely catered for by the so called ' Central ' school. Secondly, there is a large intermediate group, varying in I.Q. from about 85 to 105, who under the old régime would have been moved up into the socalled ' Senior Elementary ' schools or classes. Thirdly, there are those who need special attention and treatment, most of whom come under the official designation of ' educationally subnormal ': these would formerly have been transferred to special schools for the socalled ' mentally deficient ' or to classes for the dull and backward. As a rule in current discussion it is chiefly on the first and last groups that interest is centred; Professor Dent's contention is that the primary concern of the Secondary Modern School is the education of the big middle group.

This conclusion, as he rightly recognizes, raises several further problems: (i) how are we to decide which children fall within the great middle mass ? (ii) what standards of attainment should such children be expected to reach ? (iii) by what processes of learning can they attain them ? and (iv) what methods or techniques of teaching will best facilitate these processes ?

The first question, he maintains, can be answered according to much the same principles as those already accepted for determining the borderline at the 11+ examination. It therefore becomes the task of the teacher within the Secondary Modern School to attempt this further sub-classification. As to the second question, most teachers, he believes, would endorse Dr. Dempster's dictum that the " great middle mass are below the threshold of specialized academic or technical education ". Nevertheless, he himself maintains that, within the limits set by their innate powers, children of modest intellectual ability are capable of much higher standards than is currently supposed. With this every experienced psychologist would surely agree. As to the third problem, teachers themselves so he finds, seem to rely on one or other of two broad types of learning process: some believe that the instruction to be given should develop out of the interests of the children; others think the child's learning should be based on appropriately chosen ' special courses ': he himself, " not being a psychologist ", finds it difficult to suggest any more conclusive argument. The last question, that of methods and techniques, is in his view a matter primarily for the teachers, and, he hopes, will be progressively solved by increased experience. Already, in their attempts to frame suitable curricula and devise appropriate modes of instruction, the new ' modern ' schools had from the outset a good deal of accumulated experience to guide them; and he notes that many of their characteristic activities were in fact adapted from those developed by the more enterprising central and senior elementary schools half a century ago.

However, as Professor Dent himself has indicated, each of the four questions urgently calls for prolonged and intensive research, such as can best be undertaken by educational psychologists and by teachers who have received a training in the experimental and statistical techniques requisite for such investigations. His own conclusions are based primarily on personal impressions. The percentages and other statistical data that he has so industriously compiled from his ' sample inquiries ' and from replies to his questionnaires are intended merely to present a broad overall picture; and consequently now require separate and objective verification.

His ' interim report ' (as he modestly describes it) is admirably comprehensive and impartial, and provides by far the best survey of the subject that has hitherto been published. It should be closely studied, not only by members and officials of local education authorities throughout the country, and by every teacher whether in the ' modern ' school or out of it, but by all psychologists working in the field of education. CYRIL BURT.

Humanistic Psychology. By JOHN COHEN. Allen & Unwin, 1958. Pp. 206. 18*s*.

Professor Cohen has added his own voice to the growing chorus of critics who have of late so loudly condemned contemporary psychology for its sterility and lack of scientific rigour. " Psychology ", he says, " is the romping-ground for every extravaganza—an amalgam of humbug, platitude, piercing intuition, soaring flights of imagination, dull dogma, incisive reasoning, and sheer drivel ". Even his own contribution, he suggests, " should be read with ten grains of salt, one for each chapter ". He begins with what he describes as the oldest of intelligence tests, namely, the list of questions contained in the thirty-eighth chapter of the book of Job, printed in standard form, with spaces for the examinee to reply ' Yes ', ' No ', or ' Don't know ', according to his choice. Admittedly such questions as " Where wast thou when I laid the foundations of the earth ?" can hardly be answered with a simple negative or affirmative. Nevertheless, this " initiation ceremony " provides an effective way of demonstrating the mystery and complexity of the issues which confront the psychological investigator.

As might be expected from one trained at University College, Professor Cohen's approach follows much the same lines as those laid down by previous teachers at that College—Sully, McDougall, and Burt. Like them he insists on the unique nature of the problems presented by the very fact of consciousness. Like them, too, he registers a vigorous protest against the growing tendency to reduce psychology to a branch of the physiology of the nervous system and nothing more. As he puts it, the study of experience must start " from the inside ". His severest criticisms are reserved for the thoroughgoing behaviourists, typified by Clark Hull, with their doctrine of the mind as just a misleading name for neural mechanism. Theorists of this type, he says, " try to persuade us that all their cards are laid on the table, and seem innocently unaware that their sleeves are bulging with several others on which they unwittingly rely for playing their game ". Their trump card is the grand assumption that human actions are determined by blind causality just as rigorously as the movements of a cog in a watch or clock. Professor Cohen's own contention is (in the words of Eddington) that " whether or not causal determination is to be accepted in psychology, and where it is to be invoked, are points to be decided by a study of the mind itself, and not by a priori postulation ".

He is almost as critical of those mathematical and statistical psychologists who " take it for granted that nothing is necessary except a calculating machine ". " Man ", said Pascal, " is not a geometrical proposition. " And Professor Cohen holds up, as " a warning to sinners ", the fate of " a contemporary psychologist who wrote *cognito ergo absurdum*, and shortly afterwards published a treatise in which factor analysis is applied to the study of personality ".

By way of illustrating his own methodological approach he chooses " four fundamental themes ": first, the successive stages of emotional and social maturation; secondly, the personal significance of sensory experience, regarded as an experience developing in time; thirdly, the " individuality of thought "; and lastly certain concrete activities typical of daily life—illness, work, play, and literature. *Humanistic Psychology* is intended, not only as " an introductory text for the general reader, but also as a pronouncement on methodology for professional students of psychology, psychiatry, philosophy, and the social sciences ". Its lively and forcible style, together with the wealth of allusion and illustration drawn from poets, novelists, and theologians, ancient and contemporary, combine to offer a stimulating presentation of the ' humanist standpoint ' both for the beginner and for the expert.

J. F. HART.

A Survey of Social Conditions in England and Wales as Illustrated by Statistics. By A. M. Carr Saunders, D. Caradoc Jones, and C. A. Moser. Oxford: Clarendon Press, 1958. Pp. xxii + 302. 25*s*.

This volume, in which Dr. Moser now appears as the third collaborator, takes the place of a third edition of the *Survey of the Social Structure of England and Wales*, first published in 1927. Reference to earlier editions reveals several changes both in the subjects reviewed and in the arrangement of chapters. The previous chapters on ' Inborn Qualities ' and ' Poverty ' have been dropped; instead we have chapters on ' Fertility and Size of Family ', ' Personal Income ', and ' Personal Expenditure '. Those dealing with ' Classification by Occupation ' and ' Industrial Status and Social Class ' have been merged into one and rewritten: and four completely fresh topics have been added—' Nutrition ', ' Health ', ' Religion ', and ' Leisure '. A welcome feature consists in the numerous comparisons drawn between present conditions as indicated by the latest official figures or by data obtained from specific inquiries, and corresponding conditions as reported in earlier censuses. Thus the survey not only presents a detailed picture of the contemporary situation, but also throws light on recent trends. Throughout the book the authors' policy has been to " present the position as it is, not to analyse or explain ". Consequently almost every chapter reveals important problems in regard to which psychological study and research are badly required.

To the psychologist one of the most interesting discussions is that dealing with socio-economic classes and mobility. In the past, the majority of psychologists have been content to take either income or occupation as the best available index of social class, and to adopt the familiar fivefold classification introduced into the census of 1911. On the other hand, those who have approached the problem by factorial methods have been severely critical of such procedures. Similar objections are now raised by the contributors to the present study. " Although it is a simple matter to arrange the population by income, nevertheless ", they argue, " the resultant strata are not equivalent to social classes "; and there is in fact " no single measure of social class. " The last three chapters, dealing with Leisure, Religion, and Crime, are almost entirely new and have been contributed by Professor Carr-Saunders: like the earlier studies in these three fields by Burt and his co-workers, the tabulated data clearly demonstate disquieting trends which call urgently for further investigation.

The book as a whole forms by far the most comprehensive and impartial survey of social conditions at present available, and will provide a valuable work of reference for students of psychology, whether their field of interest is social, educational, statistical, or purely theoretical. L. J. Mitchell.

Experimental Designs. By William G. Cochran and Gertrude M. Cox. New York: John Wiley & Sons, Inc. Second Edition, 1957. Pp. xiv + 611. $12.00.

This admirable textbook is too little known to statistical psychologists. The edition of 1950 furnished a clear and useful practical account of those experimental designs which are especially useful for research in various biological fields—in fact, a handbook of procedures. For the present edition the whole has been carefully revised and expanded so as to bring the treatment up to date. Partly no doubt as a result of the growing interest in experimentation as applied to industry, the new book will be especially useful to the industrial psychologist, and to those carrying out researches on educational problems. As with the first edition, most of the illustrations are drawn from the field of agriculture, but the psychologist experienced in statistical methods should have no difficulty in adapting the procedures to his own special problems. Cyril Burt.

INDEX TO VOLUME XI

INDEX OF AUTHORS

Authors	*Titles*	*Page*
Bernyer, G.	Second Order Factors and the Organization of Cognitive Functions	19
Burt, C.	Definition and Scientific Method in Psychology	31
Burt, C.	Quantum Theory and the Principle of Indeterminacy	77
Burt, B., and Gregory, W. L.	Scientific Method in Psychology, II	105
Clarke, M. L.	Quantum Theory and the Principle of Indeterminacy	75
Conway, J..	The Inheritance of Intelligence and its Social Implications	171
Dingman, H. F.	The Relation between Coefficients of Correlation and Difficulty Factors	13
Gregory, W. L.	Factors of Neuroticism: A Reply	211
Gregory, W. L., and Burt, C.	Scientific Method in Psychology, II'	105
Hamilton, M.	An Experimental Approach to the Identification of Factors	161
Howard, M.	The Conversion of Scores to a Uniform Scale	199
Jackson, L. F.	Quantum Theory and the Principle of Indeterminacy	76
Lambert, C.	Factors of Neuroticism	210
Lawley, D. N.	Estimating in Factor Analysis	1
Phillips, J. L.	Heredity and Behaviour	208
Pilkington, G. W.	Scientific Method in Psychology, III	129
Pilliner, A. E. G.	The Rescaling of Teachers' Estimates	191
Taylor, J. G.	Scientific Method in Psychology, IV	133
Wheeler, D.	Studies in the Development of Reasoning in School Children	137
Willis, T. R.	Scientific Method in Psychology, I	97

BOOK REVIEWS

Authors	*Titles*	*Page*
Carr Saunders, A. M., Jones, D. Caradoc, and Moser, C. A.	A Survey of Social Conditions in England and Wales as Illustrated by Statistics	215
Cochran, W. G. and Cox, G. M.	Experimental Designs	215
Cohen, J.	Humanistic Psychology	214
Dent, H. C.	Secondary Modern Schools	212
Hogben, L.	Statistical Theory: The Relationship of Probability, Credibility, and Error	96
Mace, C. A. (Ed.)	British Philosophy in the Mid-Century	96
Yates, A., and Pidgeon, D. A.	Admission to Grammar Schools	94

PS | G2 - British

Vol. XI Part I May 1958

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
CYRIL BURT AND JOHN WHITFIELD

WITH THE ASSISTANCE OF
CHARLOTTE BANKS AND ALAN STUART
AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
E. A. PEEL
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
W. STEPHENSON
P. E. VERNON

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20*s. per part* (*U.S.A.* $3.50) *Subscription* 30*s.* 6*d. per volume* (*U.S.A* $4.50)

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The British Journal of Statistical Psychology is issued by the British Psychological Society. The subscription price per volume is 30*s.* net for Members of the Society and 30*s.* 6*d.* (post free) for non-members. The subscription price in the U.S.A. is $4,50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1 : non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, in May and November.

Papers for publication should be sent either to Sir CYRIL BURT, 9 Elsworthy Road, London, N.W.3., or to JOHN WHITFIELD, University College, Gower Street, London, W.C.1. Authors will receive 25 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

Books for review should be sent to the Editor, advertisements to the Secretary, British Psychological Society.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology* : subscriptions, 60*s.* per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology* : for this the subscription, £1 per annum, should be sent to Methuen & Co., Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society.

Papers for publication in *The British Journal of Psychology* should be sent to Professor James Drever, Department of Psychology, the University, Edinburgh; those for publication in *The British Journal of Medical Psychology* to Dr. Joseph Sandler, 96 Portland Place, London, W.1; those for publication in *The British Journal of Educational Psychology* to Professor P. E. Vernon, London University Institute of Education, Malet Street, London, W.C.1.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer : that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract : (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the ' Recommendations to Authors ' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21–73), or to those given in the paper on ' Setting Mathematics ' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional cost.

In future authors will receive two complete slip-proofs of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible.